

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Are warmer winters
cooling Americans'
climate-change
concerns? PAGES 317 & 357



ENDLESS SUMMER

RESEARCH ANIMALS

YEAR OF THE MONKEY

China bids for primate research leadership

PAGE 300



HISTORY OF SCIENCE

REVIEWING THE SITUATION

Two centuries of fretting about peer review

PAGE 306

EMISSIONS

AUDIT LOCAL EFFORTS

Are cities and companies meeting climate goals?

PAGE 303

NATURE.COM/NATURE

21 April 2016 £10

Vol 532, No 7599



THIS WEEK

EDITORIALS

EXPERTISE Quantum world offers politicians best of both **p.282**

WORLD VIEW How the world defines dangerous climate change **p.283**



NEW YORK Mice that make it there, can make it anywhere **p.285**

Monkeying around

China, with its freedom from the ethical pressures experienced by researchers elsewhere, is poised to become the go-to country for work on non-human primates.

In China, this is the year of the monkey. And for Chinese science, the rise of research with monkeys promises to make the next few years especially noteworthy. The country has for a couple of years been struggling to implement a major brain project, partly because officials have been trying to work out what its unique angle should be, given that similar efforts are already under way in Europe, the United States and Japan. Now researchers say that the project will be announced soon, and that primate research will feature heavily. It is natural for China to be willing to extend this line of research — and useful for the rest of the world, because elsewhere it is becoming more difficult.

As we explore in a News Feature on page 300, access to abundant non-human primates gives researchers in China a serious advantage in many key areas of preclinical study, as well as in emerging areas of basic science. With their prodigious new-found ability to apply gene-editing techniques in monkeys, for example, Chinese researchers are converting that abundance into prominent publications.

The importance of China is clear from the contrasting picture in Europe. Munich, for example, is one of Germany's main hubs for biomedical research. The large biomedical campus on the outskirts of the city is home to a large number of exciting basic-research centres and successful biotechnology companies. Last week, it announced another splendid addition — a stem-cell centre, due to open in 2017. The centre, which will provide a research and drug-screening platform based on the use of induced pluripotent stem cells, is designed to boost the translation of basic research into medical applications by identifying candidate therapies. But there is a hitch.

CRUCIAL ROLE

The German federal government, like most European governments, is pushing translational research in medicine, and its planning can't be faulted — except for one thing. It has no strategy for the primate research and testing that will be needed to move many candidate therapies into the clinic. There is a reason for that. Primate research in some German centres and elsewhere is under pressure from campaigners and politicians. The Netherlands voted late last month to revisit a 2013 parliamentary question on whether and, if so, how the country could end primate research within a decade — a stance that is sharply at odds with its own heavy promotion of translational research.

This hardening of attitudes comes as researchers and governments alike seem to agree that monkey research has a major role in many fields. That point was made dramatically during the Ebola crisis, when therapies based on monkey studies were successfully rushed into use; and ongoing investigations in Zika might make the point again.

Why does that general agreement not translate into consistent support for the primate studies that will be needed now and in the future to extend crucial areas under investigation, especially in neuroscience? The same government agencies must also consider the views of society at large, which, as our understanding of monkeys' capabilities and

commonalities with humans grows, is ever more strongly against research on primates. This is a valid perspective and needs to be balanced with the societal benefits that can be gained only through primate research. Such research is not an all-or-nothing proposal; it is one that requires continuous debate over where the research is warranted. Right now, however, too many politicians in Europe are shunning that debate, taking the easy way out and withdrawing support. This leaves a big gap in the translational-medicine programmes that they support. Researchers in China, who do not face these barriers of cost, logistics, regulations or ethical pressure, have a huge strategic advantage. This will give China's brain initiative plenty of room to thrive and leverage to negotiate research collaborations with researchers elsewhere.

Still, China needs to proceed with caution. Researchers should not just rush, for example, to be the first to tweak some gene in monkeys, even though the growing number of gene-modification tools makes it easy. That these techniques work in monkey genomes has been proved by Chinese groups already. The likely clinical and fundamental break-

"Researchers should not just rush to be the first to tweak some gene in monkeys."

throughs offered by these experiments must be well mapped out before new projects start.

This is especially true for experiments on psychiatric and neurodegenerative disorders, which, from a therapeutic perspective, offer some of the most compelling reasons to experiment with monkeys. There's no use in scientists tinkering with monkey genes to create a biomedical model if the experiment doesn't allow an investigation of how genes affect the cells, circuits and behaviours thought to be implicated in brain disorders. This requires a rigorous comparison of behaviour in the modified monkey and a normal monkey. Because the ultimate aim is to learn about human behaviour and pathology, comparison with the human disease must also be included. In many cases, these will be totally new areas of investigation not just for China, but for researchers around the globe. Few have expertise with the necessary reach. Some have experience in basic systems neuroscience, but little idea of how to approach disease. Others might work on rodent disease models, but have little understanding of primate brain function.

China should consider hiring or collaborating with scientists who are up to speed on both monkey and human behaviour. Many researchers in Europe and the United States who are keen to carry their projects through to the monkey model are sure to be interested. More-structured top-level coordination with the national brain initiatives elsewhere is also worth exploring. Chinese researchers' freedom from animal-rights pressures will probably continue for the foreseeable future, but it is not a given. To maintain that support, and to make it easier for researchers elsewhere to form collaborations, they will have to show that they are abiding by principles that guide the international scientific community — that monkeys should be used only when necessary and in as small a number as possible. ■

Red-tape tangle

Attempts by the European Union to stimulate innovation are stifled by bureaucracy.

The damning report released by auditors last week on the European Institute of Innovation and Technology (EIT) was predictable. Since it was conceived about 10 years ago, the EIT — a €3-billion (US\$3.4-billion) mechanism that is supposed to stimulate innovation in areas that are considered to be among Europe's foremost societal challenges — has suffered more than just teething problems (see page 291).

As the auditors pointed out, the EIT has struggled to align business and research communities in sectors such as public health or the development of clean technologies in a way that could address common market failures. The EIT as a whole has still to prove that its existence makes a real difference. To do so, managers must monitor more closely — and demonstrate more plausibly — whether the substantial tax money spent on the EIT triggers the desired effects on innovation.

Creating commercially relevant knowledge through basic research needs incentives. But innovation is not something that technocrats (or bureaucrats) can easily order. Innovation and bureaucracy are in fact not a good match — too much of the latter is one of the reasons why the EIT has failed to meet expectations.

The audit report comes as proposals swirl for yet another European Union innovation body — one to be called the European Innovation Council. The idea might seem inappropriate at a time when top-down approaches to stimulate absent market forces have been weighed and found wanting.

But the EIT's failure is a good occasion to think about what is missing. It's a given that the EU needs to unlock its innovative potential to make its ageing societies fit for the future and create jobs for the next generation. So why are the EU's economic competitors in North

America and Asia more able to transform the ideas of academic scientists and engineers into marketable goods and services?

It is not for want of good intent and trying. European universities have long ceased to be academic havens where students and staff ponder the wonders of the world in splendid isolation. Science parks, incubators and technology-transfer offices have become the rule on European campuses. Also, the European Commission's €80-billion Horizon 2020

“There are just too many programmes, too many levels, too many forms, bodies and exceptions.”

research programme has a strong emphasis on producing applicable science in partnership with small and large companies. Other schemes — EU Finance for Innovators, Joint Technology Initiatives, European Innovation Partnerships and the EU Innovation Union — likewise intend to obtain the maximum economic return on research money.

And yet the quality in question is in short supply. Why hasn't the investment and effort led to greater innovation?

The byzantine complexity of the EU's innovation support is making it less effective than policymakers would like it to be. There are just too many programmes, too many levels, too many forms, bodies, requirements and exceptions. The bureaucratic confusion is not stifling innovation all together — the EU's graphene flagship project and countless small entrepreneurial success stories are sufficient evidence that some things do work very well. But given the EIT dilemma, Europe's leading research universities have rightly reminded policymakers that streamlining and simplifying EU innovation instruments is a better approach to stimulating the sought-after quality than adding another layer of complexity on top of it.

This does not mean that a European Innovation Council — for which the European Commission issued a call for ideas in February — would necessarily be wasted money. But such a council must seek to optimize, rather than add to, the existing portfolio of initiatives and mechanisms. Europe's paradoxical innovation bureaucracy might still benefit from a high-level advisory body comprising competent business leaders, researchers and policy experts. So, incidentally, might the floundering EIT. ■

Expect knowledge

We are gratified when a politician shows that they know about science, but they all should.

“Swans sing before they die —” said poet Samuel Taylor Coleridge, “’Twere no bad thing/Should certain persons die before they sing.” Now, not everyone can carry a tune. Neither can everyone act any better than the average block of wood — which is why people at large seem to lend credence to singers, actors and other celebrities when they effuse on subjects that they know nothing about.

No one can doubt the prodigious acting talent of Robert De Niro, but does his turn as the young Vito Corleone in *The Godfather Part II*, or the tortured Travis Bickle in *Taxi Driver*, qualify him to opine on the link between vaccination and autism? Is he talking to me? I repeat: is he talking to me? (Clue for any readers bewildered by this: despite statements made by De Niro last week, there is no evidence for any link between vaccination and autism.)

Politician Sarah Palin has no acting ability, save that which might be parodied by the comedian Tina Fey, yet she has power and influence, which makes her increasingly barbed attacks on the reality of anthropogenic climate change all the more worrying. (Further guide for the perplexed: despite Palin's latest statements on the subject, also last week, yes, anthropogenic climate change is real.)

As the weekend approached and science had its head in its hands at the way it was being treated (again) by the news, salvation of a sort appeared. No less a person than Justin Trudeau, the debonair Prime Minister of Canada, offered an impromptu (and accurate) explanation of quantum computing at a press conference. In response, parts of the Internet have exploded into what can only be described as a nerdgasm. Why the eruption of reaction, one is entitled to ask? Shouldn't we expect all our elected representatives to be so conversant with the scientific issues of the day that explanations of quantum computing by any one of them should barely twitch a cat's whisker?

At this point, one might take a duster to scientist and novelist C. P. Snow's oft-cited 1959 tome *The Two Cultures* and refer wearily to the preponderance of a humanities education among the political class. Yet the most cursory scan of the news headlines shows how important science is to human well-being. Emerging diseases, energy policy, transport, conservation and, yes, climate change and vaccination — almost every sphere of government requires at least some familiarity with science. Especially given that most science funding is still disbursed by politicians on behalf of the public.

The problem is that science, if done properly, rarely comes up with the sound-bite certainties and expedient spin that politicians demand — nor the ability to say one thing while meaning something quite different. So perhaps it is not so surprising that the latest brave attempt by a politician to grapple with science involves the quantum world, where it is possible for something to be both true and false at the same time. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv



The Paris Agreement has solved a troubling problem

By endorsing a limit of 1.5°C, the climate negotiations have effectively defined what society considers dangerous, says **Simon L. Lewis**.

The Paris Agreement for tackling climate change opens for governments to sign this week, four months after it was agreed. The momentum created by the deal, described as a multilateral political triumph, looks set to continue: China and the United States are among the 130-odd countries expected to bring the agreement into force early by adding their signatures on the first day.

Is this the beginning of the end of the fossil-fuel age, as some suggest? It could be — its influence is certainly being felt. Peabody Energy, the largest private coal company, lost 12.6% of its value the day after the Paris deal was agreed. It filed for bankruptcy last week. But even before countries queue up to sign, the Paris Agreement could already have solved one of the most troublesome problems in the climate arena, one that has plagued scientists and policymakers for almost a quarter of a century. And yet almost nobody — scientists included — seems to have noticed.

The Paris Agreement has finally defined the threshold for ‘dangerous’ climate change. It is 1.5°C above pre-industrial levels. True, this definition is not explicitly spelled out in the agreement text. It is a de facto definition. But it is there all the same. And that is hugely significant.

Back in 1992, the world agreed to “prevent dangerous anthropogenic interference with the climate system” by signing up to the United Nations Framework Convention on Climate Change. Unfortunately, ‘dangerous’ was left undefined. Politicians asked scientists for an answer and many researchers, me included, batted the question back to them. Dangerous to whom? Climate change disproportionately affects the poor, the vulnerable and the powerless. Heatwaves tend to kill the very young and very old. Heavy storm events affect those in precarious shanty towns. Of course, if they are high enough, then emissions will affect everyone, but differentially. Even today’s global average temperature of 0.9°C above pre-industrial levels is dangerous for some, even deadly. What risks are acceptable for society? Dangerous, as many have pointed out, is a social and political question of trade-off, justice and ethics that science can inform, but not decide.

From this contested space, the figure of 2°C rose up the global agenda. It was chosen by European politicians and their advisers because it was a simple round number, seemed achievable and would prevent many catastrophic impacts. By 2010, the UN process noted a ‘shared vision’ of limiting warming to 2°C. But it was always a somewhat arbitrary choice.

At the UN negotiations in Paris in December, a more satisfying solution to the dangerous question, rather obvious in hindsight, slowly emerged. The answer was not 2°C. And it did not come from Europe.

Storm surges and ongoing sea-level rise will make low-lying small island states unviable places

to live. Some islands will cease to exist. A new grouping of 20 countries in Paris, the Climate Vulnerable Forum, painted dangerous climate change as a threat to their very existence. The rallying-cry, “1.5 to stay alive”, repeated in forcefully eloquent language in the negotiating sessions, increasingly made sense.

The Marshall Islands then deftly revealed a secret ‘high-ambition coalition’ at the talks. It included rich and poor countries alike, unravelling old geopolitical alliances, and so allowing a much more ambitious agreement to be reached.

That’s a crucial part of why China, the United States and the rest will this week sign the Paris Agreement, a UN legal instrument to hold the mean global surface-air temperature to “well below 2°C above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5°C above pre-industrial levels”.

Against the odds, vulnerable states got their message across. Without stringent limits on temperature rises, whole nations within the UN system may become stateless, which, self-evidently, is dangerous to those states.

The move is good politics, but is it based on good science? Projections of sea-level rise are notoriously uncertain, but unabated emissions rises would certainly lead to a rise of many metres over the course of a few centuries as the Greenland and Antarctic ice sheets lose mass. Such losses are nonlinear and once begun are essentially irreversible. Models now show that much of this could be avoided if emissions are curbed such that warming stays below 1.5°C. Impacts on staple crops also increase drastically after 1.5°C. Scientific evidence is on the side of the small island states.

The emergence of 1.5°C as a serious policy position comes with important lessons for scientists. The global research community has shockingly little to say on the probable impacts of a 1.5°C rise. (The Intergovernmental Panel on Climate Change last week scrambled to commission a special report on the subject.) Most impact studies and future-scenario analyses focus on 2°C and higher. Few focus on the most-vulnerable regions. It is the same bias that neglects the study of diseases that kill millions outside the developed world. Most scientists and most funds for science, after all, are from developed countries, and so tend to follow the agendas of the dominant class of those societies. In this way, science further entrenches inequality.

This bias is dangerous. And it will continue until more scientists challenge the agenda of their funders and examine their own preferences for research questions to answer. ■

Simon L. Lewis is professor of global-change science at University College London and the University of Leeds, UK.
e-mail: s.l.lewis@leeds.ac.uk

THE GLOBAL
RESEARCH
COMMUNITY HAS
SHOCKINGLY
LITTLE TO SAY ON
THE PROBABLE
IMPACTS
OF A 1.5°C RISE.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/hma2is

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

Spinning black holes align

Black holes at the cores of galaxies tend to spin in the same direction.

This unexpected result comes from a survey of 65 galaxies in the same patch of sky by the Giant Metrewave Radio Telescope near Pune in India. Andrew Russ Taylor and Preshanth Jagannathan at the University of Cape Town in South Africa measured the orientation of jets of plasma shooting out from the strong magnetic fields swirling near the supermassive black holes. They found that more than a dozen jets pointed in the same direction, suggesting that the black holes have the same spin direction. Their galaxies, which are tens of millions of light years apart, form a large 'filament' across the sky.

The galaxies may have all spawned from one enormous structure of rotating gas in the early Universe, the authors suggest.

Mon. Not. R. Astron. Soc. 459, L36–L40 (2016)

IMMUNOLOGY

Parasite promotes gut health

Infection by certain intestinal worms may help to foster the growth of bacteria that protect against gut inflammation.

Mutations in the *NOD2* gene are linked to an inflammatory bowel disease called Crohn's. In mice lacking this gene, researchers discovered that introducing parasitic helminths to the gut helped to reverse cellular defects in the small intestine and reduce inflammation.

Ken Cadwell at the New York University School of Medicine and his colleagues found that the worms promoted the growth of Clostridiales bacteria, which suppressed inflammation-causing Bacteroidales microbes in the gut. The researchers also observed a similar protective balance of intestinal microbes in people from rural areas (where helminth infections are relatively common), which was not seen in people from urban regions.

The results support the hygiene hypothesis, which says that a 'cleaner' microbial environment has made some

people more susceptible to inflammatory disorders.

Science <http://doi.org/bfc6> (2016)

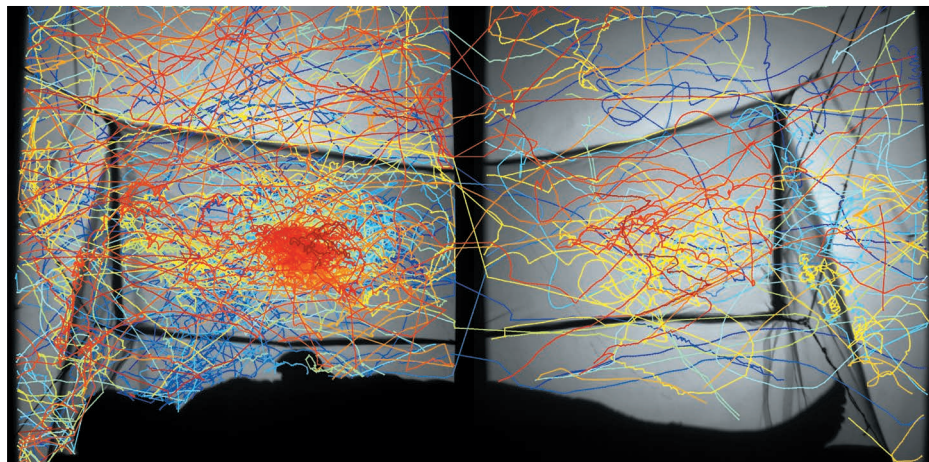
PARTICLE PHYSICS

Theories abound for new particle

Four research groups have proposed the existence of various new particles to explain an anomalous signal picked up by the two largest particle detectors at the Large Hadron Collider at CERN, Europe's particle-physics lab in Geneva, Switzerland.

Physicists with the ATLAS and CMS detectors

announced in December 2015 that they had found an excess pair of photons with a combined energy of 750 gigaelectronvolts. Christoffer Petersson at Chalmers University of Technology in Gothenburg, Sweden, and Riccardo Torre at the Swiss Federal Institute of Technology in Lausanne say that the two photons could come from the decay of a boson that is the 'super partner' of a goldstino fermion. Yuichiro Nakai at Harvard University in Cambridge, Massachusetts, and his team suggest that the new particle is made of two exotic quarks held together



BEHAVIOURAL ECOLOGY

Video reveals mosquito antics

Video monitoring has shown that mosquitoes spend most of the time near the head of a person lying under a bednet at night.

Mosquitoes carry several human diseases but are difficult to study in the field. David Towers at the University of Warwick in Coventry, UK, and his co-workers used infrared light-emitting diode backlighting to make mosquitoes visible on video. They filmed the insects with two cameras at night around a mosquito net in the laboratory and in the field in Tanzania. The team developed algorithms to track individual

mosquitoes, and found that the insects focused their efforts around the roof of the net, above the person's head (mosquito tracks pictured as coloured lines). The mosquito species that transmits West Nile virus (*Culex quinquefasciatus*) tended to be more active than the carrier of human malaria, *Anopheles gambiae*.

The technique could help to improve understanding of how mosquito behaviour affects disease transmission.

J. R. Soc. Interface 13, 20150974 (2016)

J. R. Soc. Interface/CC BY 4.0

by a force similar to the strong nuclear force.

Gang Li at Peking University in Beijing and his co-authors propose that the mystery signal is produced by a particle similar to the Higgs boson, but six times more massive. And Won Sang Cho at the Institute for Basic Science in Daejeon, South Korea, and his collaborators propose that the two photons would be only part of the debris produced by a potentially much more massive particle.

More data are needed to confirm whether this is a sign of a new particle or merely a statistical bump.

Phys. Rev. Lett. 116, 151804; 151802; 151803; 151805 (2016)

VIROLOGY

Zika protein structure solved

Scientists have elucidated the structure of a key protein that helps the Zika virus to infect cells.

George Gao and Yi Shi of the Chinese Academy of Sciences in Beijing and their colleagues crystallized the NS1 protein (**pictured**) from the Zika virus involved in the 2015 outbreak in Brazil. They report that the protein has structural similarities to those found in closely related viruses such as the dengue and West Nile viruses. But NS1 has some key differences from these related proteins — most notably, a different pattern of electric-charge distribution in a region that interacts with host cells.

The team suggests that the structural differences

could explain the diversity of symptoms caused by Zika and related viruses in humans.

Nature Struct. Mol. Biol. <http://dx.doi.org/10.1038/nsmb.3213> (2016)

MATERIALS

High efficiency for solar cell

A solar cell made of organic materials has reached record efficiency in converting sunlight into electrical energy.

Organic solar cells could be cheaper and more flexible than silicon-based ones, but they are less efficient. To boost efficiency, Jianhui Hou at the Chinese Academy of Sciences in Beijing, Feng Gao at Linköping University in Sweden and their colleagues made a device by combining the polymer PBDB-T with a small molecular compound known as ITIC. The cell, measuring 1 square centimetre, had an efficiency of 10.78% under standard conditions — a record for an organic device of this size, and close to the record of 11.5% for all polymer-based devices.

Because the cell does not use large, relatively expensive molecules known as fullerenes, it could be cheaper to produce than the current best polymer-based devices.

Adv. Mater. <http://doi.org/bd7p> (2016)

IMMUNOLOGY

Immune cells move in fast

A type of immune cell quickly moves into the liver to aid tissue repair — but the cells come from the

surrounding cavity, not from the blood.

Immune cells responding to injury have long been thought to migrate into tissues from the bloodstream. Some of these cells then mature into macrophages over 2–3 days to help heal wounds. Jing Wang and Paul Kubes at the University of Calgary in Canada used microscopes to follow macrophages over time in live mice that had sustained a liver injury. They found that mature macrophages moved into the liver from the surrounding cavity within just one hour. There, the macrophages helped to clear the injured area of dying cells so that new blood vessels could grow.

The team identified the molecular signals that drove this process, and suggest that it could occur in liver infection and in diseases such as cancer.

Cell <http://doi.org/bd8b> (2016)

EVOLUTION

Town mice differ from country ones

Urbanization can shape the evolution of the wild mouse in as little as 400 years.

Jason Munshi-South of Fordham University in New York and his colleagues analysed the genomes of 191 white-footed mice (*Peromyscus leucopus*; **pictured**) in 23 populations from both rural and urban areas in and around New York City. They conclude that the urban mice diverged from rural populations less than 400 years ago, when people began to urbanize New York. Genetically isolated mouse populations in large city parks, such as Central Park, split from other city populations around the time that the parks were first built.

Roads and buildings



obstruct most mouse migrations, making urbanization a strong enough selective pressure to affect the evolutionary history of organisms, the authors say.

Biol. Lett. 12, 20150983 (2016)

CLIMATE CHANGE

Storms intensify in the centre

Researchers studying the effects of global warming on rainstorms have found that warmer air redistributes moisture towards the middle of storms, focusing heavy rain over a smaller land area.

Ashish Sharma of the University of New South Wales in Sydney, Australia, and his colleagues analysed data from 1,300 rain gauges and 1,700 temperature stations across Australia to test how air temperature affects the intensity and spatial organization of storms. They found that atmospheric moisture was more concentrated near the storm centre in warm storms than in cooler ones, resulting in more intense peak precipitation in that area.

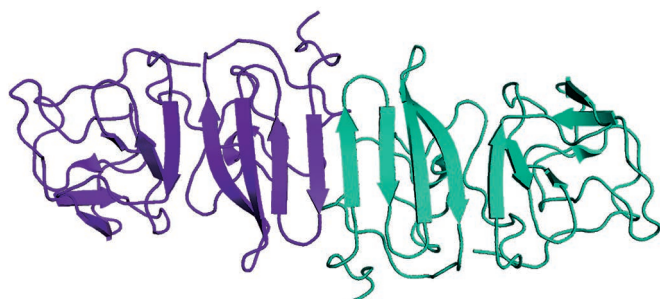
Storms that extend over a smaller area but release increasingly intense rainfall in a warming climate might boost local flood risk and severity, the authors conclude.

Geophys. Res. Lett. <http://doi.org/bd7s> (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

EVENTS

Zika link declared

The US Centers for Disease Control and Prevention (CDC) has declared that the mosquito-borne Zika virus causes microcephaly — babies born with abnormally small heads — and other fetal brain defects. The announcement, on 13 April, is based on a review of evidence by CDC researchers (see S. A. Rasmussen *et al.* <http://doi.org/bfc2>; 2016).

The mosquito season in the southern US states is looming, and the agency says that strong causal messages will reinforce advice about precautions. Some scientists caution that the proof is not yet unequivocal, but that the CDC is justified in erring on the side of caution.

Antarctic cruise

Swiss coordinators of the planned international Antarctic Circumpolar Expedition announced on 18 April the 22 scientific projects selected to take place on the research cruise. On 20 December, a 55-strong research crew involving scientists from 30 countries will set out from Cape Town on a three-month voyage on board the Russian research vessel *Akademik Treshnikov*. The scientists hope to extensively probe the Southern Ocean and map unexplored biota around Antarctica. The expedition is largely funded by the Swedish philanthropist Frederik Paulsen, founder of Ferring Pharmaceuticals.

Vaccine switch

Between 17 April and 1 May, 155 countries will introduce a new kind of polio vaccine as part of a global push to eradicate the disease. The switch will replace a 'trivalent' vaccine against the three serotypes of poliovirus with a more effective vaccine that



LUIS ACOSTA/AFP/GETTY

Earthquakes strike Ecuador and Japan

A magnitude-7.8 earthquake struck Ecuador's coast on 16 April, collapsing buildings and killing hundreds of people. The death toll was 413 as *Nature* went to press. It was the country's most powerful quake since 1979 and it devastated towns near the coast. Separately, a series of

shallow earthquakes shook Japan's Kyushu island last week, culminating in a magnitude-7 tremor on 16 April that killed at least 42 people. Buildings including a student residence, as well as turrets on a seventeenth-century castle, collapsed in Kumamoto prefecture.

targets the two types of virus that are still circulating. Just 10 cases of polio caused by a wild virus have been reported this year, in Pakistan and Afghanistan.

RESEARCH

Future health spend

The provision of aid for global health has stagnated since the 2008 financial crisis, following years of increases during the early 2000s, and international health-spending inequalities will persist as a result, predicts a report (J. L. Dieleman *et al.* *Lancet* <http://doi.org/bfdr>; 2016). In a companion report, data extrapolated from health spending between 1995 and 2013 suggest that nearly half of low- and lower-middle-income countries are likely to miss an internationally agreed goal to spend at least US\$86

per person on health by 2040 (J. L. Dieleman *et al.* *Lancet* <http://doi.org/bfdr>; 2016). By that time, the wealthiest countries will spend an average of \$9,019 per person on health, compared with \$164 per person by the poorest countries.

POLICY

Glyphosate rule

The European Parliament has called on the European Commission to restrict its marketing authority for the widely used herbicide glyphosate to seven years, amid controversy over whether the chemical may be harmful to health. The commission had instead proposed a 15-year renewal of the authority — which expires in June — to market glyphosate in European Union member states. Parliament's resolution on

13 April also calls for a new independent safety review and a restriction of glyphosate use in public areas. The resolution has no legal authority, but might influence a May vote by member states on the proposal.

CRISPR crops

The US Department of Agriculture said on 13 April that it will not regulate a mushroom genetically modified with the CRISPR-Cas9 gene-editing tool. The mushroom can now be cultivated and sold without passing through the agency's regulatory process; it is the first CRISPR-edited organism to receive a green light from the US government (see page 293). And on 18 April, DuPont Pioneer in Johnston, Iowa, announced plans to commercialize high-starch varieties of maize (corn)

BRIAN J. SKERRY/GETTY

that have been genetically modified with CRISPR to boost yields. The company aims to have the maize available within five years.

Untested drug

Brazil's President Dilma Rousseff has signed a law allowing patients to access an untested, unapproved compound that some claim is a miracle cure for cancer. The law, which went into effect on 14 April, allows those with a certificate verifying that they have cancer to obtain the drug; no prescription is required. The news came just weeks after Brazil's Ministry of Science, Technology and Innovation released laboratory results showing that the compound does not kill cancer cells grown in culture. See go.nature.com/gwzswx for more.

Whales threat

Marine-mammal experts have urged US President Barack Obama to halt permits for seismic oil and gas surveys along the mid- and southeastern US Atlantic coast. Fewer than 500 North Atlantic right whales (*Eubalaena glacialis*, pictured) remain in the wild, 27 right-whale experts from the United States, Canada and Britain said in a 14 April letter to Obama. Noise pollution from the airgun blasts used to return information about oil and



gas deposits would affect the animals on important feeding and breeding grounds, the letter says.

Warming review

The Intergovernmental Panel on Climate Change will review the possible effects on humans and ecosystems of a rise in global temperature of 1.5°C above pre-industrial levels. At a meeting on 11–13 April in Nairobi, the group agreed to produce three special reports: one looking at the impacts of 1.5°C of warming, with the other two assessing the impacts of climate change on land use and terrestrial ecosystems, and on oceans, glaciers and polar ice sheets. See go.nature.com/aq3yhf for more.

FUNDING

detector are feasible and the mission could launch sooner than planned, an expert panel reported on 12 April (see *Nature* 531, 30; 2016). The chair of the Gravitational Observatory Advisory Team, University College Dublin physicist Michael Perryman, told the BBC that the group will suggest a launch in 2029, which would bring forward the proposed start date of the €1-billion (US\$1.1-billion) mission by 5 years.

Exxon sponsorship

The board of the American Geophysical Union (AGU) has decided to continue to accept sponsorship money from the oil and gas giant ExxonMobil, despite a February letter from more than 170 AGU members and others complaining about the company's past role in spreading climate misinformation. "We

Green light

The European Space Agency's ambitious plans to build a space-based gravitational-wave

COMING UP

21 APRIL

The United Nations hosts a high-level debate on implementing its sustainable development goals for 2030, in New York. go.nature.com/ku8o5l

22 APRIL

Sentinel-1B, a radar observation satellite developed by the European Space Agency, will launch from Sinnamary, French Guiana. go.nature.com/9pmfp7

The Paris Agreement on climate change, adopted in December, will be signed in New York. go.nature.com/7fpxfw

concluded that it is not possible for us to determine unequivocally whether ExxonMobil is participating in misinformation about science currently," AGU president Margaret Leinen wrote in a blog post on 14 April describing the board's vote. Last year, the AGU accepted US\$35,000 in support from ExxonMobil.

Cancer institute

The Parker Foundation, a charity in San Francisco, California, has committed US\$250 million to harnessing the immune system to fight cancer. The money will support more than 40 laboratories at 6 centres of medical research, which together will form the Parker Institute for Cancer Immunotherapy, the foundation announced on 12 April. The institute — to be led by immunologist Jeffrey Bluestone of the University of California, San Francisco — will manage any intellectual property that emerges from the collaboration.

➔ NATURE.COM

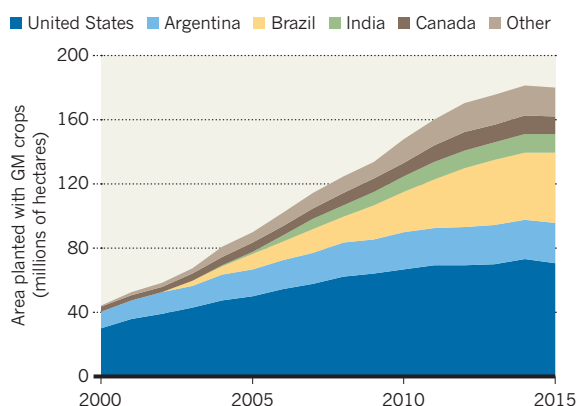
For daily news updates see: www.nature.com/news

TREND WATCH

The area planted globally with genetically modified (GM) crops declined in 2015. The 1% decline — the first in the technology's 20-year global commercial history — was primarily due to a decrease in both GM and non-GM crops caused by low prices, says the body that tracks such crops. But the International Service for the Acquisition of Agri-Biotech Applications also said in its 13 April report that major growers of GM crops, such as the United States, are approaching saturation.

GENETICALLY MODIFIED CROPS ON THE WANE

After nearly two decades of continual increase globally, the area planted with GM crops fell by 1% in 2015.



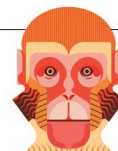
SOURCE: ISAAA

NEWS IN FOCUS

GEOLOGY Probing the underbelly of a North Korean volcano **p.290**

SNAKEBITE Synthetic biology could boost supply of antivenom **p.292**

DIRTY MICE Pet-shop rodents enhance their lab-bred brethren **p.294**



ANIMAL RESEARCH China's move to lead in monkey research **p.300**

ANNA TÄRNHUVUD



Fredrik Lanner, a stem-cell biologist at the Karolinska Institute in Stockholm, is preparing experiments that involve editing genes in human embryos.

GENETIC MODIFICATION

Embryo-editing research gathers momentum

Scientists proceed with human-genome-editing experiments as ethical debate continues.

BY EWEN CALLAWAY

At the Karolinska Institute in Stockholm, Fredrik Lanner is preparing to edit genes in human embryos. It's the kind of research that sparked an international frenzy in April last year, when a Chinese team revealed that it had done the world's first such experiments¹.

But Lanner doesn't expect his work, which will explore early human development, to cause such a fuss. A year of discussion about the ethics of embryo-editing research, and perhaps simply the passage of time, seems to have blunted its controversial edge — although

such work remains subject to the same ethical anxieties that surround other reproductive-biology experiments. "At least in the scientific community, I sense more support for basic-research applications," says Lanner, who gained approval for his experiments last June.

His instinct seems to be borne out by the fairly muted reaction to a 6 April report² of an experiment to edit human embryos — only the second to be published. A team led by Yong Fan at Guangzhou Medical University in China used the gene-editing technology CRISPR-Cas9 to try to introduce a mutation that makes humans resistant to HIV infection.

"I don't think there is anything wrong with

what these scientists have done," says Sarah Chan, a bioethicist at the University of Edinburgh, UK. "This work isn't seeking to do what is still ethically in question. It's not seeking to create genetically modified human beings."

The ethics committee of the university-affiliated hospital that approved Fan's work says that it has green-lighted two other embryo-editing projects; such research is ethically sound because it will lead to improvements in gene-editing technology and could help to prevent diseases, a committee spokesperson says.

Last December, an international summit of scientists and ethicists declared that gene editing should not be done in human embryos ▶

► that are intended for use in establishing a pregnancy — but it endorsed basic research.

“People are more understanding of this research,” says Fan, who points to UK fertility regulators’ approval in February of a proposal by developmental biologist Kathy Niakan to edit genes in healthy human embryos, at the Francis Crick Institute in London.

Fan’s team began its experiments in early 2014 and originally submitted the paper to *Cell Stem Cell*, Fan says. By the time the manuscript ended up on the desk of David Albertini, editor-in-chief of the *Journal of Assisted Reproduction and Genetics*, a different Guangzhou-based team had become the first to report human-embryo-editing experiments. That paper¹, which tried to correct a mutation that causes a blood disease, fed into a firestorm over the ethics of modifying human reproductive cells (or ‘germline’ modification). Some researchers called for a moratorium even on proof-of-principle research in non-viable embryos.

Albertini, a reproductive biologist at the University of Kansas Medical Center in Kansas City, felt that it was important to publish Fan’s paper to educate scientists and clinicians. He says that the manuscript went through two rounds of review over eight months — twice as long as is normal for the journal — and that he urged the researchers to discuss the ethical issues surrounding germline editing in the paper.

Fan’s paper should help to reassure

international observers about the legitimacy of human-embryo-editing research in China, says Robin Lovell-Badge, a developmental biologist at the Crick. More such embryo-editing papers are likely to be published, he adds. “I know that there are papers floating around in review,” he says. “I’d much rather everything was out in the open.” (Fan says that his team is now focusing on improving the efficiency of CRISPR using human stem cells).

Research involving the editing of human embryos will begin soon elsewhere in the world, if it hasn’t done so privately already.

“I know that there are papers floating around in review.”

In a *Cell* paper³ published on 7 April, Lanner’s team analysed gene expression in 88 early human embryos and is using those data to identify genes to disrupt in embryos using CRISPR–Cas9. Lanner will discuss the work at a meeting on human gene editing organized by the US National Academy of Sciences and National Academy of Medicine this month in Paris. He says that the experiments could begin in the coming months.

Norms for conducting human-embryo editing are still taking shape. Evan Snyder, a stem-cell scientist at the Sanford Burnham Prebys Medical Discovery Institute in La Jolla, California, says that whenever possible, researchers should use alternatives, such as embryos

of non-human primates. And when it is not, they should use only surplus embryos that would ordinarily be discarded from *in vitro* fertilization clinics. Both Chinese teams used non-viable embryos, but Lovell-Badge says experiments in normal embryos are also important: to see, for instance, if CRISPR–Cas9 is more or less effective in such cells.

Some scientists contend that gene-editing experiments designed to probe human development, such as those planned by Lanner and Niakan, are more valuable than experiments that are intended to lay the groundwork for creating genetically modified humans. “At the moment, there seems little point in pursuing long-term clinical goals when there’s so much not known about the technique with human embryos,” says Lovell-Badge.

But Chan thinks there should be ethical latitude for both kinds of research to proceed. “We should give the public the credit for being able to understand the difference between research into genetically modified embryos and genetically modifying human beings,” she says. “I think it’s a good thing if the hubbub dies down a bit.” ■

Additional reporting by David Cyranoski.

1. Liang, P. *et al. Protein Cell* **6**, 363–372 (2015).
2. Kang, X. *et al. J. Assist. Reprod. Genet.* <http://dx.doi.org/10.1007/s10815-016-0710-8> (2016).
3. Petropoulos, S. *et al. Cell* <http://dx.doi.org/10.1016/j.cell.2016.03.023> (2016).

SEISMOLOGY

North Korea lets scientists peer inside massive volcano

Seismic images from unprecedented international collaboration hint at future eruption risks.

BY ALEXANDRA WITZ

A rare collaboration between North Korean and Western scientists has probed the ground beneath a dangerous volcano on the Chinese–North Korean border. The work illuminates the geological plumbing that could underlie possible future eruptions.

“This is our first glimpse into what the insides of the volcano look like,” says Kayla Iacovino, a volcanologist with the US Geological Survey in Menlo Park, California.

She and her colleagues, led by Ri Kyong-Song of the Earthquake Administration in Pyongyang, North Korea, used seismic data to pinpoint molten rock beneath the mountain. The researchers’ paper was published on 15 April in *Science Advances*¹.

Called Mount Paektu on the North Korean side and Changbaishan on the Chinese side, the volcano is considered one of the region’s most hazardous. Around AD 946, it let loose one of the most powerful eruptions in recorded history, showering ash as far away as Japan. Today, more than 1.6 million people live within 100 kilometres of Paektu.

“This volcano is quiet at the moment, but it’s definitely got potential,” says team member James Hammond, a seismologist at Birkbeck, University of London. “We need to keep an eye on it.”

Lava could erupt as much as 20 kilometres away from the mountain’s summit, says Haiquan Wei, a volcanologist at the China Earthquake Administration in Beijing who has studied the mountain’s past activity².

Because the volcano straddles the Chinese–North Korean border, scientific studies have been fragmented between the two countries. “People have spent their whole lives studying the volcano and have never seen it from the other side,” says Iacovino. The mountain holds a special significance in North Korea as the purported birthplace of both the founder of the first Korean kingdom and the former North Korean leader Kim Jong-Il.

Paektu last erupted in 1903. In 2002 it began shaking, generating thousands of tiny earthquakes, possibly as molten rock shifted underground. The seismic unrest ended after several years without any lava erupting — but the episode prompted researchers on both sides of the border to reassess what they knew about the volcano and to try to prepare for



Mount Paektu (also called Changbaishan), with its crater lake, straddles the China–North Korea border.

what it might unleash in the future.

In 2011, at the invitation of the North Korean government, Hammond went to the country with Clive Oppenheimer, a volcanologist at the University of Cambridge, UK. That meeting spawned an unprecedented collaboration to try to understand Paektu better from the Korean side³. With diplomatic support from the American Association for the Advancement of Science in Washington DC and the Royal Society in London, Hammond arranged to bring six state-of-the-art seismometers into North Korea.

It wasn't easy. It took years to sort out the proper import licences, and the team had to ditch plans to measure conductivity beneath the volcano because the required equipment has a second use in submarine detection. But in the end, Hammond and his colleagues deployed the seismometers in a 60-kilometre-long line east from Paektu's summit, deep into the countryside. "Every year, I would visit these families and they would look after our stations for us," says Hammond. "They clearly wanted to understand this volcano."

The seismometers remained in place from August 2013 to August 2015 (which meant that they were not present during any of North Korea's four nuclear-weapons tests). By analysing how seismic waves travelled beneath the volcano, the scientists found that a significant part of the crust must be at least partially molten. "Whether or not that melt is going to turn into an eruption is a bigger question," says Iacovino. "But at least we can now start to draw a picture of what's happening."

Previous studies have hinted at the presence of molten rock beneath the volcano, says Haibo Zou, a geoscientist at Auburn University in Alabama. But "any new serious research", he says, "is of interest".

Chinese and North Korean scientists monitor Paektu using their own seismic networks as well as gas samples collected from hot springs. But until geologists have a better understanding of what the volcano has done in the past, it will be hard to tell emergency officials how they should prepare for future eruptions, says Iacovino.

For instance, she has been mapping the geology of the ash, pumice and other rocks thrown outward in the AD 946 eruption. Enormous clouds of superheated gas and ash swept downhill, followed by destructive mudslides.

If Paektu were to erupt again, it might send torrents of water downhill from the summit lake, or clouds of ash skyward, which could interfere with aeroplane flights across Korea and Japan.

By studying rocks collected during a 2013 visit, Iacovino has found that the AD 946 eruption probably spewed much more sulfur dioxide into the atmosphere than earlier studies found⁴. That suggests that Paektu has the potential to alter the global climate.

Hammond will be in Pyongyang this week, working on proposals to expand studies of Paektu. "We'd really like to work together with the Chinese and North Koreans to study the volcano as a whole volcano, using instruments on both sides of the border," he says. "Ultimately, it's up to them to work together, and maybe we can be a part of it." ■

1. Ri, K.-S. *et al. Sci. Adv.* **2**, e1501513 (2016).
2. Wei, H., Liu, G. & Gill, J. *Bull. Volcanol.* **75**, 706 (2013).
3. Hammond, J. 'Understanding volcanoes in isolated locations: Engaging diplomacy for science' *Sci. Diplom.* **5**, No. 1 (2016).
4. Xu, J. *et al. Geophys. Res. Lett.* **40**, 54–59 (2013).

INVESTMENT

EU innovation hub slammed

Audit picks out management and funding issues.

BY QUIRIN SCHIERMEIER

An ambitious initiative to fuel innovation in the European Union has swallowed up a vast amount of money, but has little to show for the investment, auditors say.

Backed with almost €3 billion (US\$3.4 billion) in EU cash, the European Institute of Innovation and Technology (EIT) was established in 2008 to stimulate economic growth by transferring research and innovation from academia to commercial applications. But the institute is a long way from achieving its goals, and is beset by management problems, ill-suited short-term grants and potential conflicts of interest, according to an audit report released on 14 April.

"Significant legislative and operational adjustments are required," says Alex Brenninkmeijer, the chair of the European Court of Auditors' team behind the report.

First conceived as a research powerhouse to rival the Massachusetts Institute of Technology (MIT) in Cambridge, the EIT ended up as a distributed network of academic and business partners that has failed to convince many experts. "Stimulating actual innovation by adding more bureaucracy where less is needed just doesn't work," says Helga Nowotny, a science-policy adviser to the Austrian government.

The institute operates through Knowledge Information Communities (KICs): groups of universities, research institutes and businesses working in specific technologies.

The three initial KICs set up in 2010 created 90 start-ups, 400 business ideas and 71 new or improved products, services or processes. And the EIT claims that every euro spent from its budget triggers four extra euros for innovation. But the auditors say that this is "undemonstrated and implausible" and that most of the KICs' claimed activities would have been carried out whether or not the EIT existed.

The failure to deliver has to some extent been the result of "limited leadership abilities" and high staff turnover, including at senior management level, the auditors say.

"We agree with this report's important finding and recommendations," says Martin Kern, the Budapest-based interim director of the EIT. "But we have moved on since, and we have already addressed many concerns." ■ [SEE EDITORIAL P.282](#)

See go.nature.com/vwfe2k for a longer version of this story.



Milking snakes is a key part of producing conventional antivenom.

MEDICINE

Synthetic biology tackles antivenom

Artificial antibodies could ease global snakebite burden.

BY CARRIE ARNOLD

When the medical charity Médecins Sans Frontières called the worldwide shortage of snake antivenom a public-health crisis last September, Brazilian biochemist Paulo Lee Ho wasn't surprised. He has spent his career at São Paulo's Butantan Institute searching for better ways to create antivenom to treat bites from coral snakes.

Conventional methods rely on natural coral-snake venom, which is hard to come by: the snakes produce only small amounts with each bite and are hard to raise in captivity. So Ho and others have turned to proteomics and synthetic biology in the hope of improving the quality and availability of antivenom. "We need a new way to meet the demand for antivenom from the Ministry of Health," he says.

These efforts are bearing fruit. Last month, Ho and his colleagues reported¹ that they had engineered short pieces of DNA that, when injected into mice, triggered antibodies against coral-snake venom. The scientists

then boosted the animals' immune response by injecting them with small pieces of synthetic venom antibodies synthesized in *Escherichia coli* bacteria. In a separate study², another group of researchers in Brazil used synthetic antibody fragments to neutralize the effects of bites by the pit viper *Bothrops jararacussu*.

Such progress is encouraging, given the severe medical burden caused by snakebites in the developing world, says Robert Harrison, head of the Alistair Reid Venom Research Unit at the Liverpool School of Tropical Medicine, UK. Each year, around 90,000 people die after being bitten by venomous snakes³.

Yet antivenoms are still made using a method that has not changed for more than a century. Large animals, typically horses, are injected with small amounts of purified proteins extracted from snake venom, which prompts the production of antibodies. Plasma containing these antibodies is then given to snakebite victims.

But this life-saving treatment is limited in important ways. Each antivenom is effective against only a single species or, at most, a small

group. And the drugs must be refrigerated, difficult in tropical countries without reliable electricity. "When you think about it, it's amazing these antivenoms work at all," says Leslie Boyer, director of the Venom, Immunochemistry, Pharmacology and Emergency Response Institute at the University of Arizona in Tucson.

The number of pharmaceutical companies that make antivenoms is declining, because the drugs are not very profitable. In 2010, for instance, pharmaceutical giant Sanofi of Paris ended production of the antivenom Fav-Afrique, which is designed to treat the bites of ten of Africa's most poisonous snakes.

Ho hopes that his approach will help to fill this void. Rather than relying on venom from live coral snakes, he began with small pieces of coral-snake DNA that code for venom toxins. He and his colleagues injected these DNA pieces into mice to prime their immune systems; a month later, they gave the animals a booster shot containing synthetic venom antibodies.

Only 60% of mice injected with a lethal dose of coral-snake venom survived after receiving Ho's experimental treatment, compared to nearly 100% for existing antivenoms. But Ho is undaunted. "This result shows there are other ways to obtain neutralizing antibodies," he says. "Maybe to get better results, we need to try again but use more antibodies. We just don't know yet."

The second Brazilian team, led by molecular biologist Carla Fernandes of the Oswaldo Cruz Foundation, a biomedical-research institute in Porto Velho, tested a different technique. The researchers used a phage-display library — a method of studying interactions between proteins — to make synthetic versions of antibodies that llamas produced when they were injected with *B. jararacussu* snake venom. Giving these antibodies to snakebite victims would eliminate the need to use animal plasma. It also could leave less muscle damage and tissue death at the site of the bite than conventional antivenoms, because the synthetic antibodies are smaller and better able to penetrate into tissue.

The path to newer antivenoms is not straight, but researchers think that moving quickly is key. "There has been significant, rapid progress in this area, but it needs to be fast. There are too many people dying from what is essentially a preventable disease," says Harrison.

To Boyer, however, the antivenom shortage is not caused by a lack of science. "It costs 14 bucks to make a vial of antivenom that costs \$14,000 in the US," she says. "You're not going to get cheaper than that. The expensive parts aren't the science — it's everyone wanting a cut of the profits that drives the price up and puts it out of reach." ■

1. Ramos, H. R. *et al.* *PLoS Negl. Trop. Dis.* **10**, e0004484 (2016).
2. Prado, N. D. R. *et al.* *PLoS ONE* **11**, e0151363 (2016).
3. Harrison, R. A., Hargreaves, A., Wagstaff, S. C., Faragher, B. & Lalloo, D. G. *PLoS Negl. Trop. Dis.* **3**, e569 (2009).

CAMILA CARVALHO/INST. BUTANTAN

BIOTECHNOLOGY

Gene-edited CRISPR mushroom escapes US regulation

A fungus engineered using CRISPR–Cas9 can be cultivated and sold without oversight.

BY EMILY WALTZ

The US Department of Agriculture (USDA) will not regulate a mushroom that has been genetically modified with the gene-editing tool CRISPR–Cas9, the agency has confirmed. The long-awaited decision means that the mushroom can be cultivated and sold without passing through the agency's regulatory process — making it the first CRISPR-edited organism to receive a green light from the US government.

“The research community will be very happy with the news,” says Caixia Gao, a plant biologist at the Chinese Academy of Sciences Institute of Genetics and Developmental Biology in Beijing, who was not involved in developing the mushroom. “I am confident we'll see more gene-edited crops falling outside of regulatory authority.”

Yinong Yang, a plant pathologist at Pennsylvania State University (Penn State) in University Park, engineered the fungus — the common white button mushroom (*Agaricus bisporus*) — to resist browning. The effect is achieved by targeting the family of genes that encodes polyphenol oxidase (PPO), an enzyme that causes browning. By deleting just a handful of base pairs in the mushroom's genome, Yang knocked out one of six PPO genes — reducing the enzyme's activity by 30%.

AGENCY RULES

The mushroom is one of about 30 genetically modified organisms (GMOs) to sidestep the USDA's regulatory system in the past 5 years. In each case, the agency's Animal and Plant Health Inspection Service (APHIS) has said that the organisms — mostly plants — do not qualify as something that the agency must regulate. (Once a crop passes the USDA reviews, it may still undergo a voluntary review by the US Food and Drug Administration.)

Several of the plants that bypassed the USDA were made using gene-editing techniques such as the zinc-finger nuclease (ZFN) and transcription activator-like effector nuclease (TALEN) systems. But until now, it was not clear whether the USDA would give the same pass to organisms engineered with science's hottest tool, CRISPR–Cas9.

Yang first presented the crop to a small group of USDA regulators in October 2015, after being encouraged to do so by an APHIS



STUART MCCALL/GETTY

The common white button mushroom (*Agaricus bisporus*) has been modified to resist browning.

official. “They were very excited,” Yang says: “There was certainly interest and a positive feeling” at the meetings. He followed up with an official letter of enquiry to the agency later that month.

The USDA's answer came this week. “APHIS does not consider CRISPR/Cas9-edited white button mushrooms as described in your October 30, 2015 letter to be regulated,” the agency wrote to Yang on 13 April.

“I am confident we'll see more gene-edited crops falling outside of regulatory authority.”

Such organisms were necessary for genetically modifying plants and fungi in the 1980s and 1990s, when the US government developed its framework for regulating GMOs. But newer gene-editing techniques that do not involve plant pests are quickly supplanting the old tools.

The United States is revamping its rules for regulating GMOs, which collectively are known as the Coordinated Framework for Regulation of Biotechnology. To that end, the US National Academies of Sciences, Engineering and Medicine have convened a committee that is charged with predicting what advances will be made in biotechnology products over the next five to ten years. It will hold its first meeting on 18 April.

In the meantime, Yang is mulling over whether to start a company to commercialize his modified mushroom. Fruits and vegetables that resist browning are valuable because they keep their colour longer when sliced, which lengthens their shelf life. In the past 18 months, biotech companies have commercialized genetically engineered non-browning apples and potatoes.

“I need to talk to my dean about that. We'll have to see what the university wants to do next,” says Yang about the prospect of bringing his mushroom to market. But he notes that in September 2015, Penn State filed a provisional patent application on the technology. ■



LAURENT GESLIN/NATUREPL.COM

Wild mice that have grown up exposed to a wide variety of microbes have stronger immune systems than do lab-reared mice.

IMMUNOLOGY

Dirty room-mates make lab mice more useful

Housing lab mice with pet-shop mice gives them more human-like immune systems.

BY SARA REARDON

Most scientists stocking a lab with mice order the animals from a commercial supplier. But David Masopust, an immunologist at the University of Minnesota in Minneapolis, chose a more difficult route. He decided to catch wild mice in a barn at a petting zoo — where he then had to fend off a cockatoo vying for one of the trapped mice. “You do question what exactly you got yourself into,” he says.

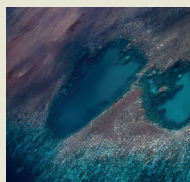
Masopust went to such lengths because he doubted that commercial lab mice — selectively bred in sanitized environments — are good research analogues for people, who do not live in such clean conditions. In a paper online in *Nature* (Beura, L. K. *et al.* <http://dx.doi.org/10.1038/nature17655>; 2016), he and his colleagues show that wild mice and mice from pet shops have strong, complex immune systems that mimic those of adult humans. Exposing lab mice to these ‘dirty mice’ strengthened their immune systems and

made them more human-like.

Therapies tested in lab mice rarely translate into humans, and researchers have long thought that differences between human and rodent immune systems may be partly to blame. Lab mice, for instance, have very low levels of some types of the immune cells called memory CD8⁺ T cells compared with adult humans. In people, these cells mature during childhood after exposure to viruses and other pathogens, and help to fight against infections and cancers.

MORE ONLINE

Q&A



Coral crisis: Great Barrier Reef bleaching is “worst we’ve seen”
go.nature.com/xamggr

MORE NEWS

- Deadly animal prion disease appears in Europe go.nature.com/rv8zvl
- First paralysed person to be ‘reanimated’ offers neuroscience insights go.nature.com/jqjnsd
- Mutant mitochondria could hinder stem-cell therapies go.nature.com/tqd5ym

NATURE PODCAST



Climate-change psychology; the 1.5°C target; and how to stop climate change ruining your work nature.com/nature/podcast

But without a better alternative, most researchers still use lab mice to test therapies or study disease. “People have known there’s a problem for a long time, but in most cases they just want to ignore it,” says Mark Davis, an immunologist at Stanford University in California. “There’s so much invested in the inbred mouse model.”

To test whether differences between lab mice and their kin are genetic or are the result of environmental exposures, Masopust housed lab mice with those bought from pet shops and monitored them for two months. The lab mice had a tough time: many became sick from exposure to the dirt and pathogens carried by the outside mice and about 20% of them died.

But the surviving lab mice emerged stronger for the experience. When the researchers infected the co-housed lab mice with *Listeria* bacteria, the animals fought off the infection much better than normal lab mice. The pet-shop mice and the co-housed mice had higher levels of mature immune cells, including CD8⁺ cells, and expressed the same immune genes as adult humans. The genes and cells of unexposed lab mice, by contrast, were similar to those of newborn humans.

“It’s so nice to see these labs starting to step out of the box,” says Amy Pedersen, an ecologist at the University of Edinburgh, UK, who studies how diseases arise in wild mice.

Pedersen is not surprised that the mice have such different immune systems. She says that researchers’ increasing interest in the microbiome — the bacteria that colonize the body — is leading to more papers supporting this idea.

Masopust says that his work should not be

“People have known there’s a problem for a long time, but in most cases they just want to ignore it.”

taken as an argument against using mice to model human diseases. Rather, he says that exposing lab mice to dirty mice could be one step in developing therapeutics. “I’d sure as heck want to pilot when possible anything I find in mice through this model before going into humans,” he says, especially because human trials are extremely expensive.

Now that he has the co-housing facility set up, Masopust hopes to collaborate with researchers who want to expose their research animals to his dirty mice. Studies on allergies, infectious disease and cancer could all be influenced by these changes to the animals’ immune systems.

Davis is impressed by the paper, calling it “a simple solution to some aspects of the differences between inbred mice and actual humans on the streets”. But he says that co-housing lab

and dirty mice will not solve all the problems with mouse models of human disease. Ideally, Davis says, mouse vendors could produce and sell lines of research mice that have been exposed to dirty mice.

But it would be logistically difficult for researchers to use these animals in clean mouse facilities, where some mice are engineered to lack immune systems, for instance, or are infected with specific pathogens for experiments. To conduct his experiments, Masopust had to convince the University of Minnesota to let him bring dirty wild mice into its meticulously clean and germ-free facilities — which required housing them in a special wing dedicated to mice with dangerous infections, such as tuberculosis.

In fact, Cory Brayton, a pathologist at Johns Hopkins University in Baltimore, Maryland, says that she is impressed that Masopust’s experiment shows just how clean typical lab mice are. “Congratulations to the mouse supplier,” she says.

Masopust says that the next phase of his research will test what happens when clean baby lab mice are raised with dirty mouse mothers. This could probe the ‘hygiene hypothesis’ — the idea that allergies and autoimmune diseases result when babies are not exposed to enough microbes or allergens when they are young. ■



ARCTIC DREAMS

Narsaq could be the site of Greenland's first major mine.

BY JULIA ROSEN

The houses of Narsaq gleam in a cheerful riot of blues, reds and yellows. The crayon-coloured town spills across a hill that separates barren mountains from a fjord filled with icebergs. But up close, grimmer details come into focus; the paint on many homes is peeling, and few signs of life stir in the narrow streets.

Established as a trading post in 1830, Narsaq long served as a hub of Greenland's fishing industry — the backbone of its economy. But in the past few decades, modernization has moved much of the fishing offshore, and many jobs in Narsaq have disappeared. The town's 1,500 residents have been struggling to find a way forward.

The same could be said of Greenland at large. Part of the kingdom of Denmark since 1814, Greenland has transformed over the past century from a society based on subsistence hunting and fishing to one built around an industrial economy and a Nordic-style welfare system. But that rapid development has stalled, leaving communities such as Narsaq to grapple with economic stagnation and high rates of unemployment. At the same time, Greenland has sought to overcome its economic and political dependence on Denmark.

"I don't know any people — any country — who don't want self-determination, who don't want independence in the world," says Hjalmar Dahl, president of the Greenlandic branch of the Inuit Circumpolar Council. Some 80% of Greenland's population is Inuit.

Over the past 35 years, Greenland has gained increasing control over its internal affairs — it was granted self rule in 2009 — but it continues to receive Danish subsidies that account for roughly one-third of its gross domestic product (GDP). To gain true independence, it will have to generate almost US\$1 billion in additional revenues — all from a population of just 56,000 people on an island with only 150 kilometres of roads and an ice sheet about 3 times the size of Texas.

But Greenlandic leaders see promise in places like Narsaq. Geological studies of the rugged peaks outside town have identified valuable deposits of rare-earth metals, uranium and zinc; a major mine is approaching the final stages of obtaining a permit. These are just some of many such deposits that have attracted the attention of international mining companies, and which proponents say could usher in a new era of prosperity.

Researchers and some residents have challenged the idea that Greenland can mine its way to independence. A bitter debate has erupted over the social and environmental impacts that mining will have on one of the last pristine parts of the planet. Now, leaders are looking for opportunities — and investors — to expand other industries such as tourism and agriculture, as well as ways to optimize Greenland's vast fishing sector. The government must juggle these goals while contending with climate change, which threatens traditional ways of life and potentially bolsters new ones.

Whatever route Greenland chooses to follow, researchers say that it needs to start paving the way now. Even if the island forgoes full political independence, Danish subsidies will remain fixed at 2009 levels, adjusted for inflation, and the funds will not help to cover the rising costs of Greenland's ageing population or to sustain small towns like

ESPER RASMUSSEN/PANOS

As it pursues independence, Greenland seeks to develop its economy without ruining one of Earth's last pristine places.

Narsaq. “It is a very urgent problem because Greenland already runs at a deficit,” says Minik Rosing, a geologist at the University of Copenhagen who is well known in Greenland for his work on the island’s future. Unless something changes, he says, “everything points toward the situation getting worse rather than better”.

RICH ROCKS

Narsaq’s name means ‘plain’ in Kalaallisut, the official language of Greenland, probably because the town occupies the flattest piece of land in sight. Mountains rise on all sides, their summits dulled by millions of years of glacial erosion. The inland ice sheet lurks just over the horizon, leaving only a thin ribbon of ice-free terrain. But what little exposed land there is happens to be rich in minerals (see ‘Mineral futures’). The crust here is ancient — up to 3.8 billion years old, in places — and has seen many cycles of volcanism and rifting. These brought metal-rich fluids close to the surface, where they formed deposits. The island also has substantial offshore oil and gas resources that could come into play if fuel prices rise or exploration costs drop.

Interest in the minerals has grown over the past decade, thanks to a confluence of forces. Greenland gained the right to manage and profit from its mineral deposits in 2009, just as the global appetite for many metals started rising. Politicians quickly pointed to mining as the best, and perhaps only, way to offset Danish subsidies and make statehood possible.

At the moment, many have their eye on the Kvanefeld deposit near Narsaq, a contender to host Greenland’s first major mine. The resource there is “potentially huge”, says Kathryn Goodenough, a geologist with the British Geological Survey in Edinburgh. She works with EURARE, an initiative to develop Europe’s rare-earth potential that brings together researchers and mining companies such as Greenland Minerals and Energy (GME), the Australian company behind the Kvanefeld project.

GME has been exploring here since 2007 and has studied core samples from hundreds of holes drilled into the nearby mountains over the years. “It’s like Swiss cheese up there,” says Ib Laursen, a company representative based in Narsaq. GME has estimated that the rocks above the town hold approximately 11 million tonnes of rare-earth oxides and that Kvanefeld is one of the largest rare-earth deposits outside China.

Another company is seeking to develop the Kringlerne deposit across the fjord, which it calls a world-class reserve of rare earths and other metals.

Until mining starts, it is not clear whether these deposits will prove as extraordinary as the companies contend, says Rosing. But the geologist, who grew up on a reindeer farm outside the Greenlandic capital, Nuuk, is optimistic about the future of the island’s mining industry. “Greenland is exceptional, it is large,” he says. “I think with enough effort, there will be definitely something happening.”

Like many Narsaq residents, Mariane Paviassen desperately hopes that the mining boom doesn’t start at Kvanefeld. She works for Air Greenland, greeting the handful of helicopter flights that touch down on Narsaq’s blustery landing pad. Her house, at the top of a narrow road on the far side of the town, is bright and inviting on a sunny day in September.

Some oppose the mine because it would bring an influx of foreign workers, but Paviassen is most worried about the uranium in the deposit, which GME plans to extract and sell along with the rare-earth elements. It’s what first brought Narsaq to the attention of scientists, including Niels Bohr, who visited in 1957 as part of Denmark’s investigations into atomic energy.

The country later banned all nuclear activity, including uranium mining, and Paviassen wishes that Greenland had upheld the tradition. “I think it is very dangerous stuff — the most dangerous stuff in the world,” she says. That’s why, in late 2014, she helped to found a citizens’ group called Urani Naamik, or No to Uranium.

GME’s current plans call for an open-pit mine on top of the plateau, about 10 kilometres from town. Paviassen’s group has highlighted the potential risks from uranium to human and environmental health, through water pollution and dust exposure. “My husband and my sons and my father — they like to go out and catch some food,” Paviassen says. But she wouldn’t eat it if mining began.

Others, including environmental organizations in Denmark, have cited the dangers of the radioactive thorium in the deposits, which currently has little commercial value, and of fluorine-containing minerals that can acidify water. Such concerns have fuelled a heated dispute over how to balance the economic benefits of exploiting Greenland’s natural resources with the environmental risks.

GME insists that Kvanefeld can be mined safely. The company says that it is considering ways to contain the thorium, and that it will lock up fluorine by converting it into a marketable mineral. “That’s a part of the demand from the government — to use best practices,” Laursen says. Studies have found¹ that modern techniques for managing tailings can minimize the contamination risk, at least in the short term. The technical details of GME’s plans, however, won’t be revealed until the Environmental Impact Assessment report comes out later this year.

Economic forces may be the biggest barrier to Greenland’s mineral plans: the prices of rare earths and other metals have slumped after reaching all-time highs in 2011. “The simple reality is, it doesn’t look good,” says Tim Boersma, a fellow at the Brookings Institution in Washington DC, who co-authored a 2014 report² on Greenland’s mining potential.

Greenland would need about 24 major mines operating simultaneously to replace the Danish subsidy, according to a 2014 joint report³ by the University of Copenhagen and the University of Greenland in Nuuk that assessed how the island’s mineral resources might shape its future. Given what is known about the deposits, that would be a tall order even in good economic times, says Rosing, who chaired the committee that wrote the report. “The dream that mining could be a quick fix for the economy — that’s not going to happen,” he says.

POWER PLANS

As the results of the report have sunk in, talk of political independence has dwindled. Many Greenlanders realize that the process will take time, and Rosing says that some young people have started to question the benefits of completely severing ties with Denmark. In their view, he says, “a nation of 56,000 people is maybe not the best way of ensuring that individuals in Greenland can shape their own future”.

Disappointments in the mining sector have spurred discussion about diversifying the strategy for economic self-sufficiency. Rosing suggests that Greenland should devise other ways to profit from what makes it unique. He is exploring the possibility of marketing rock flour — the fine powder created by glacial erosion — as a source of nutrients and neutralizing agents for tropical soils. And he says that Greenland should court industries that would benefit from its cold climate, such as computer-server farms, which use enormous amounts of energy for cooling.

Greenland has begun harnessing its torrents of glacial melt water to produce renewable energy. The island has 5 hydropower plants, and government estimates suggest that it has enough untapped potential to produce 800,000 gigawatt hours of energy per year — more than the total used by the United Kingdom and France combined. The aluminium producer Alcoa, based in New York City, has considered building a smelter to capitalize on the cheap energy and, in 2010, Greenland’s national energy company launched a pilot project using hydropower to produce clean-burning hydrogen fuel.

For the moment, however, those options are largely prospects for the future. Today, about 40% of Greenland’s workers are employed by

“The dream that mining could be a quick fix — that’s not going to happen.”

the public sector and 90% of its export economy revolves around fishing, particularly for northern shrimp (*Pandalus borealis*) and Greenland halibut (*Reinhardtius hippoglossoides*). Although catches remain good, west Greenland's shrimp stocks have declined over the past decade, perhaps influenced by climate change.

According to Helle Siegstad, director of fish and shellfish at the Greenland Institute of Natural Resources in Nuuk, the culprit could be Atlantic cod (*Gadus morhua*), a predator that could benefit from warming near Greenland and has started to reappear after being overfished. Another factor behind the shrimp's decline might be that climate change has caused a mismatch between their hatching time and the blooms of phytoplankton that they eat⁴.

But higher water temperatures have also lured new species north, such as Atlantic mackerel, Atlantic herring and even some bluefin tuna⁵, says Brian MacKenzie, a marine ecologist at the Technical University of Denmark in Kongens Lyngby. In recent years, temperatures off the coast of east Greenland have become warm enough for tuna, MacKenzie says. "It's basically a whole new habitat."

Siegstad says that the fishing fleet has been quick to pounce on these opportunities, and she is optimistic that changes in marine ecosystems will ultimately benefit Greenland's fishing industry. But even so, she worries about the island's overwhelming dependence on this variable, uncontrollable resource. "We are so sensitive," she says. "I hope we will have something else."

GROWTH INDUSTRY

Thirty minutes by boat from Narsaq, Kalista Poulsen and Agathe Devisme share 10 hectares of land with 300 head of sheep. Compared with the surrounding tundra, their grassy farm is lush. Purple wildflowers and *Angelica archangelica* — a popular medicinal herb — line their carefully manicured fields.

The couple is part of Greenland's budding agricultural industry — one of several small sectors of the economy that the island's leaders are trying to expand. Agriculture currently accounts for less than 1% of Greenland's GDP, but that figure could grow thanks to climate change, which has boosted temperatures in the south by almost 2°C over the past few decades. Modelling work⁶ by Jens Christensen and others at the Danish Meteorological Institute in Copenhagen suggests that if the world warms according to some of the most dramatic projections, the length of the growing season in southern Greenland will more than double.

But the climate is likely to become more variable, too. Already, a string of dry summers has forced farmers to import extra supplies of hay from abroad, supplementing the feed that they grow to get the animals through the long, brutal winter. This has left them wondering whether climate change will help or hurt them, says Devisme. "For the moment, it's more, kind of, disturbing."

To supplement their farming income, Devisme also runs a small bed

MINERAL FUTURES

Greenland is the world's largest island but has a population of just 56,000 people. Interest is growing in mining its rich mineral deposits, which could help Greenlanders to gain economic independence from Denmark.



and breakfast, where visitors come to relax or to fish for Arctic char in the stream behind the fields. Many Greenlanders see the island's nascent tourism industry as a welcome alternative to exploitative activities such as the mining at Kvanefjeld, which Devisme says poses a threat to her businesses.

In 2013, the government counted roughly 35,000 visitors, who contributed around 3% of GDP. Greenland hopes to ramp up adventure tourism, such as hiking and kayaking, and boost cruise-ship traffic — a pattern that has succeeded in Iceland. The consulting firm Ramboll, based in Orestad, Denmark, has projected that the tourism industry could more than double by 2025, although this would require strong investment in infrastructure such as hotels and airports, as well as increased marketing and international cooperation.

But, if Greenland is to benefit from these industries, its people must have the skills to work in them. Developing the island's human capital may be the key to Greenland's success, according to a 2013 report⁷ by the Copenhagen Institute for Futures Studies.

Today, although many Greenlanders possess a wealth of informal knowledge, only 35% of students go beyond compulsory school, which they finish at age 15

or 16. The government aims to boost the number continuing with their training, and the plan starts with strengthening elementary education.

The residents of Narsaq are doing their part. Here, late on a Sunday afternoon, workers bustle around a fenced-off construction site in the centre of town. A crane swings overhead, hoisting wooden beams onto a tower of scaffolding, where crews are renovating the red-panelled school. It will soon boast a wall of windows, giving Narsaq's children a grand view of the mineral-rich mountains, the ice-choked fjord and their own small town — as it lurches forward into Greenland's uncertain future. ■

Julia Rosen is a journalist in Portland, Oregon.

1. Committee on Uranium Mining in Virginia, Committee on Earth Resources, Board on Earth Sciences and Resources, National Research Council *Uranium Mining in Virginia* (National Academies Press, 2012); available at <http://doi.org/bd5g>
2. Boersma, T. & Foley, K. *The Greenland Gold Rush: Promise and Pitfalls of Greenland's Energy and Mineral Resources* (Brookings Institution, 2014); available at go.nature.com/14rlhh
3. The Committee for Greenlandic Mineral Resources to the Benefit of Society *To The Benefit of Greenland* (Univ. Copenhagen and Univ. Greenland, 2014); available at go.nature.com/mzqpaz
4. Koeller, P. et al. *Science* **324**, 791–793 (2009).
5. MacKenzie, B. R., Payne, M. R., Boje, J., Høyer, J. L. & Siegstad, H. *Glob. Change Biol.* **20**, 2484–2491 (2014).
6. Christensen, J. H., Olesen, M., Boberg, F., Stendel, M. & Koldtoft, I. *Future Climate Change in Greenland: Kujalleq Municipality* (Danish Meteorological Institute, 2016); available at go.nature.com/m9jybw (in Danish).
7. Copenhagen Institute for Futures Studies *Future Scenarios for Greenland* (Copenhagen Institute for Futures Studies, 2013); available at go.nature.com/rtyjs8



MONKEY KINGDOM

China is positioning itself as a world leader in primate research.

BY DAVID CYRANOSKI

An hour's drive from Kunming in southwestern China, past red clay embankments and sprawling forests, lies an unusual zoo. Inside the gated compound is a quiet, idyllic campus; a series of grey, cement animal houses stack up on the lush hillside, each with a clear plastic roof to let in the light. This is the Yunnan Key Laboratory of Primate Biomedical Research, and its inhabitants are some 1,500 monkeys, all bred for research.

The serenity of the facility belies the bustle of activity within. Since it opened in 2011, this place has quickly become a Mecca for cutting-edge primate research, producing valuable disease models and

seminal publications that have made its director, Ji Weizhi, a sought-after collaborator. Its campus houses a collection of gene-edited monkeys that serve as models of Duchenne muscular dystrophy, autism and Parkinson's disease. Ji plans to double the number of group leaders working there from 10 to 20 in the next 3 years, and to seek more international collaborations — he already works with scientists in Europe and the United States. “In terms of a technology platform, Ji is just way ahead,” says one collaborator, cardiologist Kenneth Chien at the Karolinska Institute in Stockholm.

Ji is not alone in his ambitions for monkey research. With support from central and local governments, high-tech primate facilities have sprung up in Shenzhen, Hangzhou, Suzhou and Guangzhou over the past decade. Last month, the science ministry approved the launch of a facility at the Kunming Institute of Zoology that is expected to cost millions of dollars to build. These centres can provide scientists with monkeys in large numbers, and offer high-quality animal care and cutting-edge equipment with little red tape. A major brain project, expected to be announced in China soon, will focus much of its efforts on using monkeys to study disease.

The enthusiasm stands in stark contrast to the climate in the West, where non-human-primate research is increasingly stymied by a tangle of regulatory hurdles, financial constraints and bioethical opposition. Between 2008 and 2011, the number of monkeys used in research in Europe declined by 28%, and some researchers have stopped trying to do such work in the West.

Many have since sought refuge for their experiments in China by securing collaborators or setting up their own laboratories there. Some of the Chinese centres are even advertising themselves as primate-research hubs where scientists can fly in to take advantage of the latest tools, such as gene editing and advanced imaging. “It could be like CERN in Switzerland, where they set up a large facility and then people come from all over the world to get data,” says Stefan Treue, a neuroscientist who heads the German Primate Center in Göttingen, Germany.

With China fast becoming a global centre for primate research, some scientists fear that it could hasten the atrophy of such science in the West and lead to a near monopoly, in which researchers become over-reliant on one country for essential disease research and drug testing. “Governments and politicians don't see this, but we face a huge risk,” says Erwan Bezard, who researches Parkinson's disease at the French national biomedical research institute INSERM in Bordeaux, and has set up his own primate-research company, Motac, in Beijing. Europe and the United States still have the lead in primate research, he says, but this could change as expertise migrates eastwards. “China will become the place where all therapeutic strategies will have to be validated. Do we want that? Or do we want to stay in control?”

SIMIAN SIMILARITIES

For decades, researchers have relied on monkeys to shed light on brain function and brain disease because of their similarity to humans. Growth in neuroscience research has increased demand, and although high costs and long reproductive cycles have limited the use of these animals in the past, new reproductive technologies and genetic-engineering techniques such as CRISPR–Cas9 are helping researchers to overcome these drawbacks, making monkeys a more efficient experimental tool.

China has an abundance of macaques — the mainstay of non-human-primate scientific research. Although the population of wild rhesus macaques (*Macaca mulatta*) has declined, the number of farmed animals has risen. According to data from the Chinese State Forestry Administration, the number of businesses breeding macaques for laboratory use rose from 10 to 34 between 2004 and 2013, and the quota of

animals that those companies could sell in China or overseas jumped from 9,868 to 35,385 over that time. Farm populations of marmosets, another popular research animal, are also on the rise.

Most monkeys are shipped to pharmaceutical companies or researchers elsewhere in the world, but the growing appreciation among scientists of monkey models has prompted investment by local governments and private companies in dedicated research colonies. The country's 2011 five-year plan singled out primate disease models as a national goal; the science ministry followed up by pumping 25 million yuan (US\$3.9 million) into the endeavour in 2014.

Scientists visiting China are generally pleased with the care given to animals in these facilities, most of which have, or are trying to get, the gold-standard recognition of animal care — accreditation by AAALAC International.

Ji's Yunnan Key Laboratory is the most active primate facility, but others are giving it competition. The new monkey facility at the Kunming Institute of Zoology was funded as part of the national development scheme for big science equipment that includes telescopes and supercomputers. The money will help the institute to double its colony of 2,500 cynomolgus monkeys (*Macaca fascicularis*) and rhesus macaques.

Zhao Xudong, who runs the primate-research facility, says that the plan is to “set it up like a hospital, with separate departments for surgery, genetics and imaging”, and a conveyor belt to move monkeys between departments. There will be systems for measuring body temperature, heart rate and other physiological data, all to analyse the characteristics, or ‘phenotypes’, of animals, many of which will have had genes altered. “We are calling it the ‘genotype versus phenotype analyser,’” says Zhao. It will take ten years to finish, but he hopes to begin building this year and to start research within three. Other facilities, although smaller, are also expanding and diversifying. The Institute of Neuroscience in Shanghai plans to increase its population of 600 Old World monkeys to 800 next year and expand its 300-strong marmoset colony.

A QUESTION OF COST

Outside China, the numbers are heading in the opposite direction. Harvard Medical School closed its affiliated primate facility in May 2015 for ‘strategic’ reasons. Last December, the US National Institutes of Health decided to phase out non-human-primate experiments at one of its labs and subsequently announced that it would review all non-human-primate research that it funds. In Europe, researchers say, the climate is also growing colder for such research.

Costs are a major disincentive. In 2008, Li Xiao-Jiang, a geneticist at Emory University in Atlanta, Georgia, helped to create the world's first transgenic monkey model of Huntington's disease¹ with colleagues at Yerkes National Primate Research Centre. But Li says that it costs \$6,000 to buy a monkey in the United States, and \$20 per day to keep it, whereas the corresponding figures in China are \$1,000 and \$5 per day. “Because the cost is higher, you have to write a bigger grant, and then the bar will be higher when they judge it,” says Li. Funding agencies “really do not encourage large-animal research”.

For Li, the solution was simple: go to China. He now has a joint position at the Institute of Genetics and Developmental Biology in Beijing, where he has access to around 3,000 cynomolgus monkeys at a farm in Guangzhou and some 400 rhesus monkeys at the Chinese Academy of Medical Sciences' monkey facility in Beijing. He has churned out a series of publications on monkeys with modified versions of the genes involved in Duchenne muscular dystrophy² and Parkinson's disease³.

Neuroscientist Anna Wang Roe says that red tape drove her to China. Roe's team at Vanderbilt University in Nashville, Tennessee, is

“CHINA WILL BECOME THE PLACE WHERE ALL THERAPEUTIC STRATEGIES HAVE TO BE VALIDATED.”

attempting to work out how modules in the brain are connected, and she estimates that she and her colleagues have spent 25% of their time and a good deal of cash documenting the dosage and delivery-method for each drug they administered to their monkeys, as required by regulations. “We record something every 15 minutes,” she says. “It’s not that it’s wrong. It’s just enormously time-consuming.”

In 2013, impressed by the collaborative atmosphere at Zhejiang University in Hangzhou, she proposed that it build a neuroscience institute. The next day the university agreed, and she soon had a \$25-million, 5-year budget. “Once the decision is made, you can start writing cheques,” she says. She is now closing her US laboratory to be the director of the Zhejiang Interdisciplinary Institute of Neuroscience and Technology, where she hopes to open a suite of the latest brain-analysis tools, including a powerful new 7-tesla functional magnetic resonance imaging device that she says will give images of the primate brain at unprecedented resolution.

Bob Desimone was similarly impressed with the speed at which China moves. As a neuroscientist who heads the McGovern Institute for Brain Research at the Massachusetts Institute of Technology in Cambridge, in January 2014, he had a ‘meet and greet’ with the mayor of Shenzhen. In March, the mayor donated a building on the Shenzhen Institute of Advanced Technology campus for a monkey-research facility, and the centre’s soon-to-be director, Liping Zhang, promised that it would be ready by summer. Thinking that impossible, Desimone bet two bottles of China’s prized mind-numbing liquor, *maotai*, that it wouldn’t be done in time. He lost. The group raised most of the \$10 million needed from city development grants, along with a small input from McGovern, and soon the first animals were being installed in the Brain Cognition and Brain Disorder Research Institute. “This place just makes things happen quickly,” Desimone says.

But money and monkeys alone are not enough to lead to discovery. Researchers say that China is short on talented scientists to take advantage of the opportunities provided by animal research. That’s why the organizers of the country’s new primate centres hope to attract an influx of foreigners to permanent posts or as collaborators. So far, many of those moving to China have been Chinese or foreigners with a previous connection to the country, but others are expressing interest, says neuroscientist Guoping Feng, also at the McGovern Institute. Already, the Shenzhen primate centre has recruited from Europe and the United States, and Desimone says that it will be “an open technology base. Any-one who wants to work with monkeys can come.”

EDITED MONKEYS

The rapid spread of CRISPR–Cas9 and TALEN gene-editing tools is likely to accelerate demand for monkey research: they are turning the genetic modification of monkeys from a laborious and expensive task into a relatively quick, straightforward one. Unlike engineered mice, which can be bred and sent around the world, “monkeys are difficult to send, so it will be easier for the PI or postdoc to go there,” says Treue.

Already, competition is fierce as researchers are racing for the low-hanging fruit — engineering genes with established roles in human disease or development. Almost all reports of gene-edited monkeys produced with these techniques have come from China. Desimone predicts that the pursuit of monkey disease models “could give China a unique niche to occupy in neuroscience”.

The cages of Ji’s facility are already full of the products of gene editing. One troop of animals has had a mutation genetically engineered into the *MECP2* gene, which has been identified as the culprit in humans with Rett’s syndrome, an autism spectrum disorder. An animal sits listless and unresponsive, holding tight to the bars of the cage as her normal twin sister crawls all over her. In another cage, a monkey with the mutation pumps its arm, reminiscent of repetitive behaviour seen in the

human disorder. Some incessantly suck their thumbs. “I’ve never seen that in a monkey before — never so constant,” says Ji.

Among the range of other disease models in Ji’s menagerie are monkey versions of cardiovascular disease, which he is working on in collaboration with the Karolinska Institute. And last year, Ji made the world’s first chimeric monkeys using embryonic stem cells⁴, an advance that could make the production of genetically modified animals even easier. The question now is whether these genetically modified monkeys will propel understanding of human brain function and dysfunction to a higher level. “You can’t just knock out one gene and be sure you’ll have human-like disease phenotype,” says Ji.

Researchers see an opportunity to understand human evolution as well as disease. Su Bing, a geneticist at the Kunming Institute of Zoology, is working with Ji to engineer monkeys that carry the human version of a gene called *SRGAP2*, which is thought to endow the human brain with processing power by allowing the growth of connections between neurons. Su also plans to use CRISPR–Cas9 to introduce human versions of *MCPH1*, a gene related to brain size, and the human *FOXP2* gene, which is thought to give humans unique language ability. “I don’t think the monkey will all of a sudden start speaking, but will have some behavioural change,” predicts Su.

INTERNATIONAL DIVIDE

Although the opportunities are great, there are still obstacles for scientists who choose to locate their animal research in China. Trying to keep a foot in two places can be challenging, says Grégoire Courtine, a spinal-cord-injury researcher based at the Swiss Federal Institute of Technology in Lausanne, who travels almost monthly to China to pursue his monkey research at Motac. He has even flown to Beijing, done a couple of operations on his experimental monkeys, then returned that night. “I’m 40 years old, I have energy in my body. But you need to really will it,” he says.

Another downside, says Li, is that policies can change suddenly in China. “There is uncertainty. That makes us hesitate to commit,” says Li, who has retained his post at Emory University. And the immunity that China’s primate researchers have had to animal-rights activism could start to erode, warns Deborah Cao, who researches law at Griffith University in Brisbane, Australia, and last year published a book on the use of animals in China⁵. People are starting to use Chinese social-media sites to voice outrage at the abuse of animals, Cao says.

China has competition in its bid to dominate primate research, too. Japan has launched its own brain project focused on the marmoset as a model: the animal reaches sexual maturity in a year and a half, less than half the time it takes a macaque. Some research facilities in China are now building marmoset research colonies — but Japan is considered to be several years ahead.

And some researchers want to ensure that such work continues outside Asia. Courtine says that he’s “fighting to keep alive” a monkey-research programme he has in Fribourg, Switzerland, because he thinks it’s important to have a division of labour. “Research that requires quantity, I’ll do in China. I would like to do sophisticated work in Fribourg,” he says.

Back at his primate centre in Yunnan, Ji is sure that such work is already taking place. His dream, he says is “to have an animal like a tool” for biomedical discovery. He knows there is a lot of competition in this field, especially in China. But he feels confident: “The field is wide, and there are many, many projects we can do.” ■

SEE EDITORIAL P.281

David Cyranoski writes for Nature from Shanghai, China.

1. Yang, S.-H. *et al. Nature* **453**, 921–924 (2008).
2. Chen, Y. *et al. Hum. Mol. Genet.* **24**, 3764–3774 (2015).
3. Niu, Y. *et al. Hum. Mol. Genet.* **24**, 2308–2317 (2014).
4. Chen, Y. *et al. Cell Stem Cell* **17**, 116–124 (2015).
5. Cao, D. *Animals in China: Law and Society* (Palgrave, 2015).

COMMENT

HISTORY Peer review's origins illuminate current tensions **p.306**



ARCHAEOLOGY What motivated Ice Age artists is a fraught question **p.310**

MEMOIR Hope Jahren's hymn to a research career and all its strangeness **p.312**

OBITUARY Geoffrey Eglinton, molecular-fossil pioneer, remembered **p.314**

CHEN-MIN CHUNG/IN PICTURES/COREIS



BYD, a Chinese car manufacturer, has joined other firms in committing to increasing the sales of electric vehicles by 30% by 2030.

Track climate pledges of cities and companies

Data transparency is key to accounting for how local governments and the private sector are contributing to global emissions reduction, say **Angel Hsu** and colleagues.

As world leaders agreed to cut global greenhouse-gas emissions at the climate talks in Paris last December, so too did many heads of cities, states, regional governments and companies. More than 400 mayors attended the talks, and Google, Microsoft and Adobe joined scores of large companies committed to getting all of their electricity from renewable sources.

But these pledges are largely not legally binding. What assurance do we have that chief executives and mayors will implement

their initiatives? And how can we gauge whether the sum of these efforts will prevent the global temperature from warming by 2°C above pre-industrial levels?

Tracking and verifying climate actions — commitments made by parties to prevent climate change — requires transparent data reporting. These actions can relate to climate mitigation, adaptation or financing: for instance, Coca-Cola has pledged to set an internal carbon price by 2017; the UK city of Aberdeen has committed to reducing

community carbon dioxide emissions by 42% between 2008 and 2020.

Visible results motivate participants and help independent third parties to track their progress. But much of the story is not being told. Registries for reporting climate initiatives are proliferating — but in a haphazard way. Most have a narrow focus and differ in their criteria for inclusion and reporting.

In 2014, the United Nations Framework Convention on Climate Change (UNFCCC) launched the Non-State Actor Zone for

► Climate Action (NAZCA) to compile the most ambitious and transparent commitments from different data platforms. NAZCA is by far the most comprehensive registry of climate actions made below the national level. Already, its business participants account for one-third of the global economy.

Our analysis of the data posted on NAZCA as of April 2016 shows that the platform lacks pledges from many developing countries and from heavily polluting sectors such as the fossil-fuel industry (see 'Partial view'). Small- and medium-sized enterprises and those with limited resources are deterred from declaring their efforts.

NAZCA offers a solid foundation for building a more complete and accurate platform for collating non-state emissions-reduction efforts. To realize its potential, the UNFCCC must fill the gaps by coordinating data collection and promoting the platform's use. NAZCA should widen and strengthen its criteria to encourage leaders to set universal reporting standards and ensure that local climate actions align with national goals to avoid double counting and omissions.

PATCHY REPORTING

NAZCA currently contains more than 11,300 commitments. Most are from local governments and large businesses in the developed world, especially Europe. The database relies on networks of cities and companies to share their members' data. It captures just a fraction of what is being done.

NAZCA records two types of commitment: individual and cooperative. Each is tagged in one or more sector of climate action, such as 'renewable energy', 'private finance' or 'resilience'.

We analysed only individual commitments and excluded cooperative ones because of concerns about double counting (many commitments are listed twice, once as individual and once as cooperative), which left 7,069 actions. We retained all the tags given to these commitments; actions tagged with both 'emissions reduction' and 'energy efficiency', for example, are counted in both sectors.

Our analysis revealed that NAZCA's biggest blind spots are in developing countries, including China and nations in Africa and southeast Asia. These omissions do not necessarily indicate climate inaction; many cities, states, regions and enterprises are taking actions but lack resources or the motivation to record them.

The costs of recording data vary among sectors. For example, the costs of monitoring transportation and energy use differ, and depend on access to technologies and human resources. Whereas some platforms are free to join, others are not; the Climate Registry, a North American initiative, costs between US\$750 and \$12,000 per year.

Many large companies in the developing

world are overlooked. The sole Chinese bank on NAZCA is the Agricultural Bank of China — one of the world's five largest banks in terms of market capitalization. China's other four top banks disclose climate actions on their websites or in reports, with varying detail. Twenty major Chinese banks are part of an alliance that restricts loans to high-polluting, high-energy-consuming enterprises — by 2014 it had prevented 400 million tonnes of carbon emissions. But these banks are unable to declare 'green credit' initiatives in NAZCA because there is no such category.

Many big emitters are also absent. Only 17 of the 90 fossil-fuel and cement companies responsible for about two-thirds of global historical greenhouse-gas emissions¹ report climate commitments in NAZCA. None of the oil giants from OPEC, the Organization of the Petroleum Exporting Countries, is on the roster; also missing are energy titans Coal India and Lukoil, Russia's second largest oil producer. Some omissions stem from differing corporate cultures — a US oil company may experience greater domestic pressure to include a climate-change pillar in its corporate strategies than, say, a Saudi Arabian one. Other companies may simply be unaware of NAZCA.

Several US fossil-fuel companies, including Peabody Energy, Arch Coal and Massey Energy, are not listed on NAZCA. Their vulnerability to the risks of a changing economic and regulatory climate might be their greatest incentive to participate. Faced with competition from natural gas and tighter regulation, Peabody Energy, the world's largest private coal producer and distributor — and number 16 out of the top 90 fossil-fuel and cement historical carbon emitters¹ — filed for bankruptcy in April. And Arch Coal joined several other large US coal companies in filing for bankruptcy this year.

Registering entire sectors in one go would broaden the representation of heavy-polluting industries, including those from developing countries. This approach has proved successful in the cement sector, which accounts for 5% of global carbon emissions. Companies responsible for 30% of the world's cement production participate in the Cement Sustainability Initiative, a 17-year-old programme that reports to NAZCA. The initiative's members collectively aim to reduce 2020 carbon emissions by 50 million to 100 million tonnes², equivalent to removing 10.5 million cars from the road for a year.

INCONSISTENT METRICS

Incomplete and incompatible data-collection methods thwart attempts to measure and compare mitigation efforts. Without

clear requirements, participants report data unevenly or idiosyncratically, often leaving out information on the sources and sectors targeted and the scope of particular initiatives. For instance, less than one-sixth of NAZCA's commitments related to carbon pricing include a specific carbon price, and the amounts listed range from \$0.01 to \$357.37 to emit 1 tonne of CO₂ equivalent.

Fifteen of the banks that have recorded climate actions on NAZCA, including investments in renewables and emissions reductions in their own operations, are also among the world's 20 biggest financial backers of coal mining and coal-fired electricity generation — the most carbon-intensive way to produce power³.

Essential contextual information is often absent. Urban and regional boundaries are frequently defined by local governments in contradictory ways. Political language obfuscates: distinctions between emissions from the 'government' (typically a municipality's building and transport emissions) and the 'community' (an entire community's emissions, stemming from any sources that a municipal government would exercise influence over) shift from place to place.

Sectors often lack definition. Many municipal action plans focus on sectors such as waste, transport, buildings and energy efficiency that generate high emissions and come under local control. The Covenant of Mayors for Climate and Energy, for example, is a network of 6,600 mainly European cities committed to emissions reduction. It requires its members to address at least two of four key sectors, such as emissions from residential buildings and transport. When such details are unavailable, analysts (ourselves included) must search for the information elsewhere — in published studies or in interviews — or use approximations and proxy variables. This makes analyses slow to perform, hindering the capacity of third parties to act as watchdogs.

NAZCA does not enable analysts to assess the quality of commitment data to evaluate progress. Problems arise in data collection, information deposition and in the structure of the system. For example, few data providers collect information on how or whether projects are being implemented; NAZCA does not track such data; and there is no harmonized definition of what should be tracked.

When different bodies post data, double counting of emissions reductions — in local reports and again in national ones — can be a problem for analysts trying to add everything up. For example, the United States' second biennial report⁴ to the UNFCCC includes a 'high abatement scenario' that recognizes that state-level policies may exceed federal requirements. The solution is to ensure that sub-national efforts are accounted for, or tracked separately within national climate

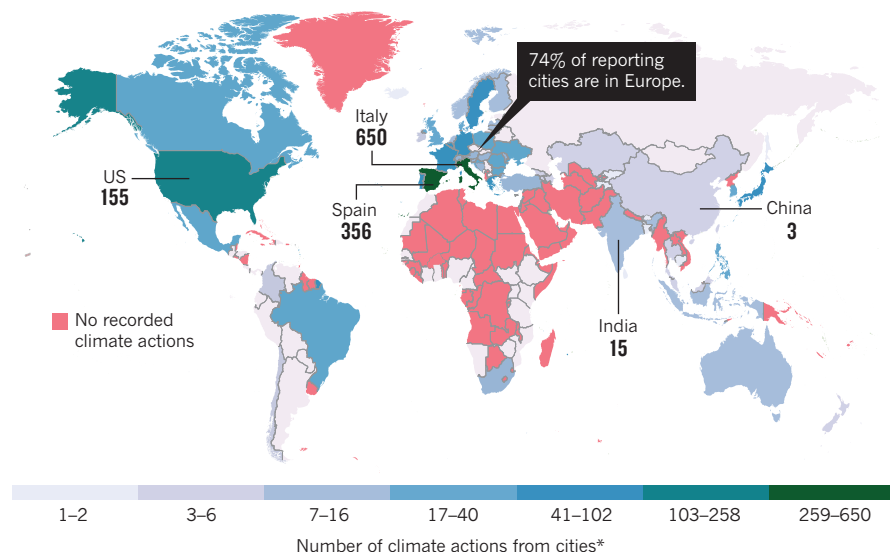
"Essential contextual information is often absent."

PARTIAL VIEW

Cities, regions, investors and civil-society organizations, mostly in the developed world (1), have so far listed more than 7,000 individual climate commitments spanning many sectors (2) in the United Nations' Non-State Actor Zone for Climate Action (NAZCA) database. Analysis of these actions highlights gaps and trends, and the need to firm up reporting of climate pledges made below the national level.

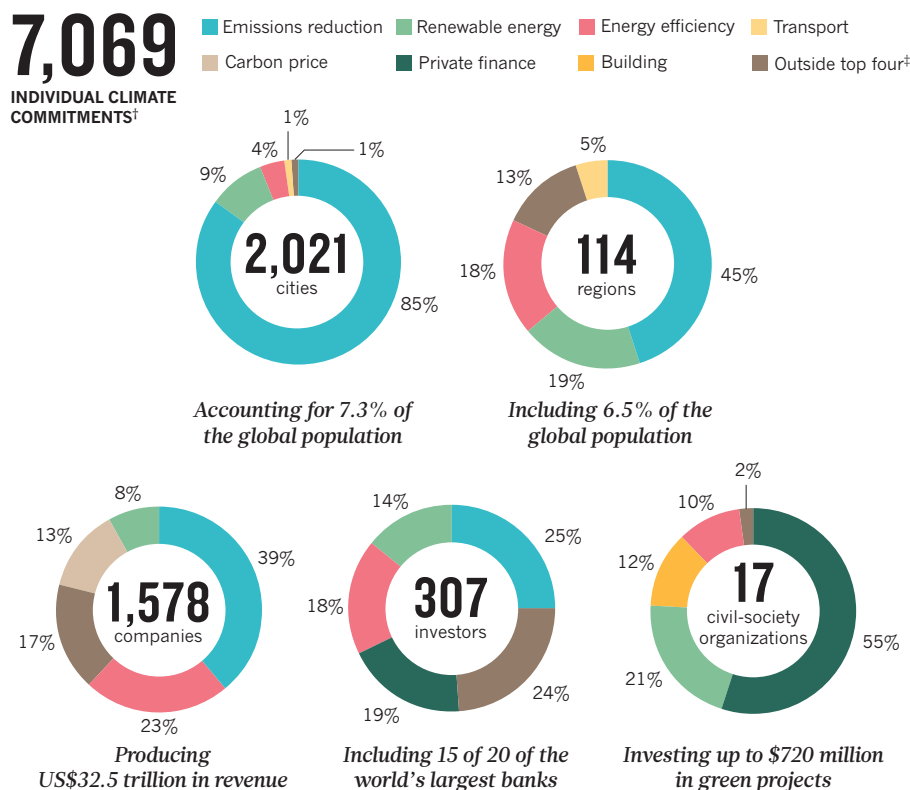
1 City initiatives

Few cities in the the developing world report their climate projects in NAZCA.



2 Differing approaches

Long-term emissions reductions are the main focus of individual cities and regions; private investors target many different short-term projects across sectors. For each, the top four sectors of climate action are shown.



action plans. Different levels of government must share their goals and harmonize their reporting. The efforts of mayors and governors who deliver mitigation results beyond national requirements need to be recognized to encourage others.

Cities and sub-national governments are limited in their jurisdiction and often short of resources. National governments can help to support local initiatives. The German government, for instance, gave the state of Saxony-Anhalt subsidies to build its renewable-energy sector⁵. Sub-national governments have more flexibility to experiment with potentially risky policy tools. Seven Chinese cities and provinces are piloting emissions-trading schemes ahead of a national programme rollout planned for 2017.

TRACK OUTCOMES

NAZCA is a good first step in understanding the global extent of climate actions by regional and municipal governments and the corporate sector. The next step is to track the outcomes of these initiatives. NAZCA and its data providers should identify common benchmarks for tracking performance and implementation. These should determine the impact and benefits (such as lower health-care costs from reduced air pollution) of climate actions, as well as the gaps in them.

Data collection is expensive. Monitoring the implementation of the UN's Sustainable Development Goals, a broad set of development targets, will require \$1 billion in annual aid to support data collection in developing countries alone⁶. Increased financial support and knowledge sharing between governments and organizations, as well as research to find cheaper ways to monitor data (avoiding overlaps and inefficiencies), would help developing countries to join in. Adding a finance 'matchmaking' function — linking funders with those who need funds — to NAZCA would attract new initiatives and ensure that existing ones receive enough backing. The Covenant of Mayors, for example, links its members to European Union public funds.

A common set of standards that embrace open-data principles needs to be developed by third-party organizations and adopted by climate-action networks. For example, a partnership of think tanks and business associations (the World Resources Institute, the C40 cities network and ICLEI—Local Governments for Sustainability) produced the Global Protocol for Community-Scale Greenhouse Gas Emission Inventories.

Cities, states, companies, investors and civil-society organizations need to use comparable reference points, time frames, levels of emissions coverage, planning processes, methodological assumptions and

*Includes both individual and cooperative actions registered in NAZCA. †Number removes double-counted or repeated actions, and excludes cooperative actions. ‡All NAZCA categories: energy access and efficiency, renewable energy, private finance, resilience, transport, building, forest, short-term pollutants, innovation, agriculture, emissions reduction, use of carbon price, other. See go.nature.com/hmzg26.

Analysis of data present in NAZCA database on 12 April 2016. To explore the data further, see visuals.data-driven.yale.edu/climateaction.

approaches. They should report a comprehensive range of emissions sources and sinks. The Paris agreement asks nations to provide this information. Ideally, carbon balance sheets should address consumption-based emissions and account for the leakage or export of emissions (when entities transfer production of emissions to other locations through trade, for example), estimated at the national level to be three times the physical quantity of traded goods⁷. Some local governments, such as that of King County in Washington state, which includes Seattle, are starting to incorporate these considerations into their commitments.

Adding a function to NAZCA that evaluates the degree of implementation for each climate action would help to compare and identify actions that are being carried out and those that exist only on paper. (This could use the ranking system of the Climate Action Tracker, which rates national climate pledges as 'inadequate', 'medium', 'sufficient' or 'role model'.) Data sets on countries' populations, gross domestic products and land areas, for instance from the World Bank, could be linked with NAZCA data to provide visualizations and metrics to understand the breadth and scope of global climate actions.

By promoting meaningful and accurate record keeping, NAZCA could become the gold standard for climate-action reporting networks. ■

Angel Hsu is assistant professor at Yale-NUS College and Yale School of Forestry and Environmental Studies, Singapore. **Yaping Cheng** and **Amy Weinfurter** are research associates and **Kaiyang Xu** and **Cameron Yick** are research assistants at the Yale Data-Driven Environmental Solutions Group, New Haven, Connecticut, USA. e-mail: angel.hsu@yale.edu

1. Heede, R. *Clim. Change* **122**, 229–241 (2014).
2. Ecosys & University of Cambridge Institute for Sustainability Leadership. *Better Partnerships: Understanding and Increasing the Impact of Private Sector Cooperative Initiatives* (2015).
3. Schücking, H., Kroll, L., Louvel, Y. & Richter, R. *Bankrolling Climate Change* (urgewald, groundWork, Earthlife Africa Johannesburg & BankTrack, 2011); available at <http://go.nature.com/wpqvqz>
4. US Department of State. *2016 Second Biennial Report of the United States of America* (US Department of State, 2015); available at <http://go.nature.com/y7dom8>
5. Hoffmann, M. J. *Climate Governance at the Crossroads* (Oxford Univ. Press, 2011).
6. Espey, J. et al. *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development* (Sustainable Development Solutions Network, 2015); available at <http://go.nature.com/oqk4yd>
7. Wiedmann, T. O. et al. *Proc. Natl Acad. Sci. USA* **112**, 6271–6276 (2015).



William Whewell, peer-review pioneer.

PAUL D. STEWART/SPL

Troubled from the start

Pivotal moments in the history of academic refereeing have occurred at times when the public status of science was being renegotiated, explains **Alex Csiszar**.

Referees are overworked. The problem of bias is intractable. The referee system has broken down and become an obstacle to scientific progress. Traditional refereeing is an antiquated form that might have been good for science in the past but it's high time to put it out of its misery.

What is this familiar litany? It is a list of grievances aired by scientists a century ago. If complaining about the faults of referee systems is nothing new, such systems are not as old as historical accounts often claim. Investigators of nature communicated their findings without scientific referees for centuries. Deciding whom and what to trust usually depended on personal knowledge among close-knit groups of researchers. (Many might argue it still does.)

The first referee systems that we would recognize as such were set in place by English scientific societies in the early nineteenth century. But these referees were never intended to play the part of supreme scientific gatekeepers. That notion emerged in around 1900 (see 'Past notes'). It was exactly then that some began to wonder whether referee systems might be fundamentally flawed. In this sense, peer review has always been broken.

Today, with the debate about the future of peer review more fraught than ever, it is crucial to understand the youth of this institution. What's more, its workings and its imagined goals have evolved continually, and its current tensions bear the marks of this. The referee system has become a mishmash of practices, functions and values. But one thing

stands out: pivotal moments in the history of peer review have occurred when the public status of science was being renegotiated.

SCIENTIFIC PUBLICISTS

In 1831, William Whewell, a Cambridge professor and philosopher of science, proposed a scheme to the Royal Society of London. He suggested that it commission reports on all papers sent for publication in the semi-annual *Philosophical Transactions*. Written by teams of eminent scholars, these reports might, he argued, be “often more interesting than the memoirs themselves” and thus a great source of publicity for science¹. Besides, authors would be grateful to know that their papers would be read carefully by at least two or three people. The society was just then launching a new journal to be called the *Proceedings of the Royal Society*, a cheaper monthly periodical to include abstracts of papers presented at the society. It had pages to fill and seemed the ideal place for these new reports.

At the time, editors of scientific journals made publishing decisions by personal fiat, perhaps in consultation with some trusted helpers. For publications that belonged to a scientific academy or society — such as the *Philosophical Transactions* — the vote of some committee of eminent persons would determine a manuscript's fate. (The temptation to conflate these practices with modern referee systems has led to the stubborn myth that the origins of the scientific referee can be traced back as far as the seventeenth century.)

Whewell was not much concerned about preventing shoddy papers from being printed; he was not proposing a new mechanism to inform publishing decisions. Instead, he was one of many people campaigning to increase the public visibility of science and give a unified identity to the scientific enterprise in England. (It was he who, a few years later, coined the word ‘scientist’ to this end.) This movement had begun in 1830 and is now most remembered for Charles Babbage's *Reflections on the Decline of Science in England*, a screed about the paucity of state funding for, and public recognition of, science. But its more consequential legacy is the referee system.

Whewell was cribbing from a century-old custom at the French Academy of Sciences in Paris of writing reports that evaluated inventions and discoveries in the service of the king. There, researchers who were elected to the academy were paid by the state as a reward for scientific eminence, and politicians seemed to value their opinions. Indeed, to be an expert (a French word not yet common in English) was almost by definition to be a writer of reports. Whewell reckoned that those French *académiciens* must be doing something right.

The proposal to turn the Royal Society into a corps of expert judges in the

style of the French academy was met with enthusiasm. But translating the report-writing practice across the Channel proved more complicated than Whewell expected.

NEWS OR VIEWS?

Whewell agreed to write the first report. His collaborator was a former student at Cambridge, John William Lubbock, a mathematically inclined astronomer who was also the Royal Society's treasurer. They jointly selected a manuscript submitted by George Airy, another up-and-coming astronomer. The paper, ‘On an inequality of Long Period in the Motions of the Earth and Venus’, used sophisticated mathematical methods to calculate how the orbits of these planets were influenced by the gravitational force each exerted on the other.

Whewell and Lubbock took turns reading the manuscript — copying technologies at the time left much to be desired. Both instantly knew what they thought of it. And they completely disagreed.

They argued about the paper for months. Both wrote draft reports, which could not have been more different. Whewell's focused on the significance of the problem and on Airy's remarkable conclusions. Lubbock's picked at the inelegant ways in which Airy had constructed his equations. Most fundamentally, they argued about what a reader's report ought to be. Whewell wanted to spread word of the discovery and to place it in the bigger picture (think *Nature's* News & Views and *Science's* Perspectives). “I do not think the office of reporters ought to be to criticize particular passages of a paper but to shew its place,” he told Lubbock. If they picked out flaws, he warned, authors would be put off. Lubbock had other priorities: “I do not see how we can pass over grievous errors,” he wrote.

Feeling that they had reached an impasse, Lubbock went to the author himself to deliver his suggestions for improvement. Airy was understandably irritated that his manuscript was being subjected to this strange new procedure. “There the paper is,” he wrote to Whewell, “and I am willing to let my credit rest on it.” He had no intention of changing his text. Lubbock threatened to pull out, but ultimately relented and swallowed his criticisms, acknowledging that this was “the first report which the Council have ever made” and trying to see the bigger picture. He thanked Whewell for putting his “shoulder to the wheel” and signed his name to the report².

With disaster averted, Whewell's version of the report was read publicly at the society on 29 March 1832, and was printed in the *Proceedings*, while Airy's full paper appeared in the *Transactions*. Lubbock's critiques never became public.

Not long before, the Astronomical Society of London (now the Royal Astronomical



1665 Henry Oldenburg, secretary of the Royal Society in London, creates the *Philosophical Transactions* to simplify his correspondence. He uses no referee system.

1699 France's Royal Academy of Sciences is given power by Louis XIV (pictured centre, with academy members) to report on and approve books for publication and bypass the royal censors.

1752 After vicious satires of the *Philosophical Transactions*, the Royal Society establishes a committee to vote on what to publish.

1831 Cambridge professor William Whewell convinces the Royal Society to commission public reports on manuscripts. Might referees increase the visibility of science?

1833 By now the reports have become private and anonymous.

1892 A pamphlet ‘On the Organisation of Science’ published in London by ‘A Free Lance’ kick-starts a movement to standardize the selection and distribution of scientific papers. Might referees be guardians of the literature?

1892 A paper surfaces that was rejected by a Royal Society referee in 1845, outlining the kinetic theory of gases more than a decade before James Clerk Maxwell's famous paper. Might referee systems be fundamentally flawed?

1968 British physicist John Ziman describes the referee as “the lynchpin about which the whole business of Science is pivoted”. Outside the United Kingdom and North America, many editors and scientists remain largely unconvinced.

1973 External refereeing becomes a requirement for publication in *Nature*¹⁰.

1991 An e-mail/FTP server at xxx.lanl.gov for freely sharing unreviewed physics preprints goes live. Later relocated to the web at arXiv.org, it becomes a touchstone for discussions about the end of peer-reviewed journals.

2006 *PLoS ONE* launches as an open-access journal that eschews ‘importance’ as a factor in peer review.

2007–11 *EMBO Journal*, the *Frontiers* series and *BMJ Open*, among other journals, experiment with open peer review, publishing reviewers' names or notes alongside papers.

► Society) and the Geological Society of London had also begun to experiment with similar reports. It was a geologist, George Greenough, who introduced the term ‘referee’ in 1817, importing into science a term he knew from his days as a law student³. But it was the Royal Society’s system of reports that caused the British scientific world to take notice. The practice gradually spread to other societies, including the Royal Society of Edinburgh and the Linnean Society of London. But it was not really until the twentieth century that journals unaffiliated with any society slowly followed suit.

ANONYMOUS JUDGES

The struggle between Whewell and Lubbock represented two distinct visions of what a referee might be. Whewell was the authoritative generalist, glancing down on the landscape of knowledge. He was unconcerned with — and probably not in a position to critique — the details. Such referees were, according to the Royal Society’s president, “Elevated by their character and reputation above the influence of personal feelings of rivalry or petty jealousy”⁴. Lubbock was a younger specialist, Airy’s equal. This allowed him to take a fine-tooth comb to Airy’s arguments; it also put him in the position of reviewing a direct competitor.

Initially, Whewell’s vision won out. But the system began to transform even as it lurched into existence. After a couple of years, the reports became shrouded in secrecy. The last *Proceedings* issue to include one was in mid-1833, and no negative reports were ever published. A letter Whewell wrote in 1836 shows that he himself had changed his view: he describes the referee as a defender of a society’s reputation, working behind the scenes to exclude publications that do not belong. Neither the Royal Society’s archives — nor the personal papers of those involved — are clear on how this happened, but we should not be surprised that it did. In England, unlike France, there was little precedent for public authorities judging from on high what constituted good or bad science. Signing one’s name to explicit criticism of a colleague would have been ungentlemanly.

More familiar was the anonymous critic who purported to speak for the public, epitomized by the anonymous book reviews that dominated English periodicals throughout the period, from the *Quarterly Review* to the lowly *Mechanics’ Magazine* (the practice survives today in *The Economist*). Through anonymity, as one uncredited editor argued in 1833, “the individual is merged in the court which he represents, and he speaks not in his own name, but *ex cathedra* (with full authority)”⁵. Justifications of the anonymity of the scientific referee took a similar view.

It took just a decade for the referee to become an established scientific persona,

and not a noble one. An 1845 exposé in a London magazine painted a picture of referees as scheming judges quite possibly “full of envy, hatred, malice, and all uncharitableness”. Hidden away in some secret chamber, this scientific judiciary, the article implied, used the cover of anonymity to advance their personal interests — perhaps through undetectable acts of piracy — at the expense of helpless authors⁶.

It was only near the turn of the twentieth century that the idea began to take hold that editors and referees, taken as one large machinery of judgement, ought to ensure the integrity of the scientific literature as a whole. Amid calls to curtail the “veritable sewage thrown into the pure stream of science” (a suggestion⁷ by the physiologist Michael Foster in 1894), English scientific societies debated combining their publishing apparatuses, with a standardized referee system overseeing all of scientific publishing. (The plan was abandoned, in part because it would have meant convincing publishers of independent journals, such as the *Philosophical Magazine*, to go out of business.)

Nonetheless, the referee was gradually reimagined as a sort of universal gatekeeper with a duty to science. As this idea gained ground, many began to worry that the system itself might be intrinsically flawed, a force that impeded creative science and which ought to be abolished. Such worries culminated in what was surely the first formal inquiry into the workings of referee systems — in 1903, by the Geological Society of London. The inquiry found that opinion was sharply divided on the subject, receiving several vitriolic statements about the injustices and inefficiencies of the systems in use. The ‘referee’ was in such disrepute that they nearly banned the use of the term in all society business.

But referee systems survived, and were slowly set up by independent journals as well. Outside the Anglophone scientific world, referee systems remained rare. Albert Einstein, for example, was shocked when an American journal sent a paper of his to a referee in 1932. The idea that any legitimate scientific journal ought to implement a formal referee system began to take hold in the decades following the Second World War.

APOTHEOSIS AND FALL

In the 1960s, refereeing emerged as a symbol of objective judgement and consensus in science. The referee was, in the words of the physicist and science writer John Ziman, “the lynchpin about which the whole business of Science is pivoted”⁸. Just as in 1830s England, the relationship of science to the public was at the foreground of these

changes. The scientific community was once again working hard to solidify perceptions of its role in society. The very phrase ‘scientific community’ dates from this time. Researchers wanted to preserve autonomy while holding on to the massive government funding that had come their way since the Second World War. Allocations for basic research in the United States, for instance, swelled by a factor of 25 in less than a decade⁹.

‘Peer review’ was a term borrowed from the procedures that government agencies used to decide who would receive financial support for scientific and medical research. When ‘referee systems’ turned into ‘peer review’, the process became a mighty public symbol of the claim that these powerful and expensive investigators of the natural world had procedures for regulating themselves and for producing consensus, even though some observers quietly wondered whether scientific referees were up to this grand calling.

Current attempts to reimagine peer review rightly debate the psychology of bias, the problem of objectivity, and the ability to gauge reliability and importance, but they rarely consider the multilayered history of this institution. Peer review did not develop simply out of scientists’ need to trust one another’s research. It was also a response to political demands for public accountability. To understand that other practices of scientific judgement were once in place ought to be a part of any responsible attempt to chart a future path. The imagined functions of this institution are in flux, but they were never as fixed as many believe. ■

Alex Csizsar is associate professor of the history of science at Harvard University, Cambridge, Massachusetts, USA.
e-mail: acsizsar@fas.harvard.edu
Twitter @alexcizsar

1. W. Whewell to P. M. Roget, 22 March 1831; Royal Society of London Library [DM/1].
2. J. W. Lubbock to W. Whewell, 27 January 1832; Trinity College Library, Cambridge [a/216/61].
3. George Greenough Papers; University College London [Add. 7918/1621].
4. *Proc. R. Soc. Lond.* **3**, 140–155 (1832).
5. *New Monthly Magazine* **39**, 2–6 (1833).
6. *Wade’s London Rev.* **1**, 351–369 (1845).
7. *Nature* **49**, 563 (1894).
8. Ziman, J. *Public Knowledge: An Essay Concerning the Social Dimension of Science* (Cambridge Univ. Press, 1968).
9. Kaiser, D. *Nature* **505**, 153–155 (2014).
10. Baldwin, M. *Making Nature: The History of a Scientific Journal* (Univ. Chicago Press, 2015).

CORRECTION

In the Comment ‘Cracking the Indus script’ (A. Robinson *Nature* **526**, 499–501; 2015), the sentence about Bryan Wells’s estimate of the number of Indus script signs has been revised to more accurately reflect Wells’s contributions. He estimated the number of signs at 676, not 958.



Aboriginal rock paintings at the Nanguluwur shelter in Kakadu National Park, Australia.

PALAEOLITHIC ART

Old masters, early cultures

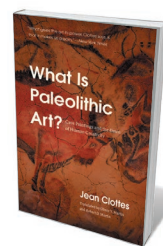
Jill Cook examines archaeologist Jean Clottes's theories about what motivated Ice Age artists.

Subtle, imaginative and brilliantly accomplished, the images of animals and humans found in caves and dated from the end of the last Ice Age, between 40,000 and 10,000 years ago, continue to astonish us. The emotions and motives that inspired them beg to be understood. In *What Is Paleolithic Art?*, Jean Clottes, the renowned cave- and rock-art specialist, suggests some answers.

After 40 years leading research at French sites such as the Chauvet and Cosquer caves and the Volp Caverns, Clottes describes his investigations into why people adorned the walls of caves and open-air rock surfaces with engravings, paintings and sculptures. His approach is challenged by some specialists, who feel that it is futile to attempt to interpret or take the investigation of cave art beyond the collation of facts that, in their view, cannot be explained except in terms of what, when, where and how.

Clottes is prepared to go further. He wants to test hypotheses developed to parsimoniously explain the majority of observed facts against the premise that the images indicate spiritual behaviours. He goes so far as to call humans *Homo spiritualis*, owing to our capacity to symbolize ideas beyond words and transcend realities through belief. Instead of viewing these ancient cave and rock artists as apart from us, he makes the case for using historical evidence to “detect convergences in ways of thinking or conceiving of particular aspects of reality”. This is not simplistic analogy, but reasoning “based on the behaviour of analogous societies”.

During the 1990s, Clottes widened his interests from European Palaeolithic sites to the rock art of other continents. At this time, he worked closely with archaeologist and ethnologist David Lewis Williams, an expert on the southern African San people and their paintings. Together, they developed a hypothesis that could account for many attributes of Pleistocene and historical rock and cave art around the world. They realized that for many traditional, non-literate peoples, the landscapes in which they live are considered to be imbued with spiritual powers; particular places, sometimes decorated with images of animals and/or symbols, are reservoirs of supernatural powers, where shamans might perform rituals to



JEAN CLOTES

What Is Paleolithic Art?: Cave Paintings and the Dawn of Human Creativity

JEAN CLOTES
(TRANSLATED BY
OLIVER Y. MARTIN AND
ROBERT D. MARTIN)
University of Chicago
Press: 2016.

➔ [NATURE.COM](https://www.nature.com)

For more on science
in culture see:
[nature.com/
booksandarts](https://www.nature.com/booksandarts)

contact, appease or exploit the spirits.

Clottes provides an overview of some of the varied beliefs and practices that he has researched, observed or been told about on visits to sites in Africa, Asia, Australasia and the Americas. He sees these “multiple realities” as part of the broad spectrum of spirituality focused in the landscape and nature. He shares his own experiences of fascinating but brief encounters with indigenous guides, such as Clifford, a venerated medicine man of the Native American Ute people. When they visited a rock-art site in Utah, a female big-horn sheep appeared. Clottes describes how Clifford chanted and made tobacco offerings, and how he later said that the sheep was “the spirit of the site” and suggested that the spot would be good for “vision-questing”.

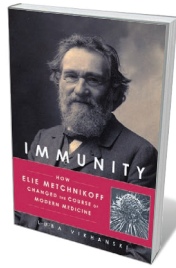
Finally, Clottes shows how such anthropological insights enrich our ability to question the Palaeolithic record and construct interpretations of behaviours, actions and events in the deep past with a better-informed historical imagination. His view is that knowledge of present and historical rock-art practices can be the key to interpreting the past. Older, less rigorous applications of this thinking were rejected for being simplistic. Clottes’s approach is more cautious, and he readily admits that the significance of many more-recent rock-art sites may be unknown or reinterpreted by modern aboriginal descendants.

Some readers will wonder whether it is right to view the varied social behaviours that resulted in drawing, painting and sculpture solely through the lens of belief. Others may question a search for universal traits that brings together eclectic cross-cultural similarities in spiritual practice that are fascinating but not necessarily helpful. Above all, Palaeolithic art is the first visible sign of modern human consciousness, of self-awareness, complex language, the use of metaphor and symbol, a sense of beauty, minds powered by brains like our own. Clottes offers little thought on how we might research these aspects of our deep history and what makes us human. But as neuroscience advances, this must surely be the next step.

This is a thought-provoking book about complex societies that endeavoured to understand the world in their own various ways. For anyone interested in Ice Age art, Clottes’s enthusiasm cannot fail to energize, inspire and provide caution to their own investigations. ■

Jill Cook is acting keeper in the Department of Britain, Europe and Prehistory at the British Museum in London. Specializing in research on portable art made of bone, antler, ivory and stone, she curated the 2013 exhibition *Ice Age Art: Arrival of the Modern Mind*, and has written a book of the same name. e-mail: jcook@britishmuseum.org

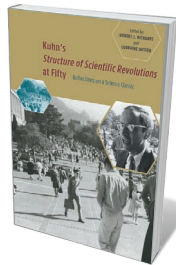
Books in brief



Immunity: How Elie Metchnikoff Changed the Course of Modern Medicine

Luba Vikhanski CHICAGO REVIEW PRESS (2016)

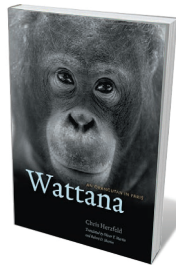
In 1882, Russian zoologist Elie Metchnikoff discovered the mechanics of natural immunity while experimenting on starfish larvae. Physician Jules Rochard called his theory an “oriental fairy tale”, yet it won Metchnikoff a share in a Nobel prize decades later. As journalist Luba Vikhanski reveals in this engrossing scientific biography, the gifted Metchnikoff’s breakthroughs extended to invertebrate embryology, treatments for syphilis and pioneering research on some of today’s hottest topics in biology: the microbiome, probiotics and longevity.



Kuhn's Structure of Scientific Revolutions at Fifty

Edited by Robert J. Richards and Lorraine Daston UNIVERSITY OF CHICAGO PRESS (2016)

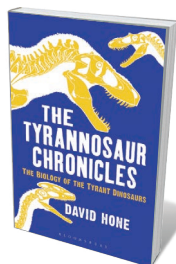
Few books leave a wake like physicist-turned-historian Thomas Kuhn’s *The Structure of Scientific Revolutions* (University of Chicago Press, 1962; see D. Kaiser *Nature* **484**, 164–166; 2012). These essays on that classic, edited by science historians Robert Richards and Lorraine Daston, emanate from a 2012 commemorative conference. It’s a scholarly treat, from George Reisch probing the cold-war roots of Kuhn’s provocations on dogma, to David Kaiser tracing the experimental psychology in his philosophical claims.



Wattana: An Orangutan in Paris

Chris Herzfeld, translated by Oliver Y. Martin and Robert D. Martin UNIVERSITY OF CHICAGO PRESS (2016)

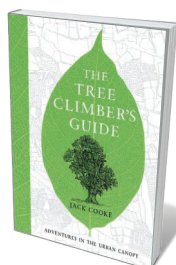
For this thoughtful, unusual study of the human–ape ‘interface’, philosopher of science Chris Herzfeld focuses on a captive orang-utan, one of less than 1,000 worldwide. Zoo-born Wattana, given string, cloth and paper at the Jardin des Plantes menagerie in Paris, made elaborate knots and ‘necklaces’ — a skilful use of fibre unsurprising in a tree-dwelling primate that builds complex nests, yet so far seen only in captivity. A trove of gripping research, somewhat marred by its scattershot presentation.



The Tyrannosaur Chronicles: The Biology of the Tyrant Dinosaurs

David Hone BLOOMSBURY SIGMA (2016)

Weighing some 6 tonnes and sporting ranks of formidable serrated teeth, the tyrannosaur has bitten into the human imagination like no other creature from the Cretaceous period ending 66 million years ago. Luckily for its devotees, the fevered pace of palaeontology means that findings on 30 or so species of tyrant dinosaur are piling up fast. Ecologist David Hone’s primer lays out the facts and educated guesses, from the beasts’ near-global distribution to their taxonomy, anatomy, reproduction, behaviour and spectacular looks, possibly enhanced by a “mosaic” of scales, feathers and keratin folds.



The Tree Climber's Guide: Adventures in the Urban Canopy

Jack Cooke HARPERCOLLINS (2016)

Nature writing has focused mainly on the terrestrial, and sometimes the marine. With Jack Cooke’s guide, it just got arboreal. An ebullient tree-climber, evoking the hominin connection to canopy life even as he mourns humanity’s lack of “biaxial ball-and-socket wrist joints”, Cooke exults in the sensory thrill of being aloft. But his tour of London’s trees, from cedar of Lebanon to common lime, is less about racing to the top than remapping the city to reconnect us with the extraordinary worlds above our heads. **Barbara Kiser**



Hope Jahren has been outspoken about the rights of women in academia.

AUTOBIOGRAPHY

A lab of one's own

Jennifer Rohn delights in the discovery-laden memoir of palaeobiologist Hope Jahren's life scientific.

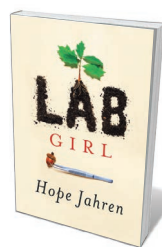
Ours is a profession still largely unfamiliar in popular culture. Scientists' memoirs are scarce, and even rarer are those by women, who are still finding their way in a male-dominated environment.

Hope Jahren is the voice that science has been waiting for. Her autobiography, *Lab Girl*, is a tell-all that demystifies a research career even as it reveals its — at times overpowering — strangeness. Jahren, a palaeobiologist at the University of Hawaii at Manoa, writes a popular candid blog (www.hopejahrensurecanwrite.com), mostly about the plight of women in research academia. But *Lab Girl* is much more. From childhood origins as a loner hiding in the classroom lab of her science-teacher father, through a 20-year career, Jahren's voice is clear, compelling and uncompromisingly honest.

Jahren studies the chemical composition of both modern and fossilized life to understand Earth's changing environment. Plants are central to this story, and twining through her account like tenacious ivy are weird and wonderful strategies that they deploy to battle a hostile world. Jahren's first

'eureka moment' is discovering that the hackberry tree (*Celtis occidentalis*) toughens its seeds by coalescing minerals into opal. We learn how trees survive winters at -40°C : their cells secrete pure water while concentrating sugars and proteins as a syrupy antifreeze, so that ice crystals do not form inside the cell. And when it comes to their progeny, the odds are vanishingly small that any one seed will reach adulthood. Plant development, from tentative rootlets to mature fruit, with plenty of knots in between, becomes a metaphor for Jahren's own progress in the challenging landscape of academia.

Much of Jahren's research takes place outdoors, and it has ranged all over the globe. Although her staple technique is state-of-the-art isotope analysis, her investigations are guided by careful observation of nature



Lab Girl
HOPE JAHREN
Knopf: 2016.

in situ. She's the type of scientist who cheerfully spends three seasons drilling 30 metres through Arctic turf, one painstaking centimetre at a time. Even the most casual observation, such as how a species of moss springs back when she removes her foot, can suggest a hypothesis. Between sessions of hard graft, her lab group takes road trips to see bizarre attractions, or attempts elaborate campfire cuisine.

Amid descriptions of Jahren's work, uncomfortable secrets of science are laid bare. Jahren breezily reveals how scientists sometimes work on tangents not mentioned in their grants. She confesses that writing papers can be an exercise in narrative retrofitting, a "size-zero mannequin designed to showcase the glory of a dress that would be much less perfect on any real person". The sterile language of papers also obliterates the human drama, such as when data must be reprocessed after a student quits, not wanting to consign herself to a life like Jahren's.

Above all, Jahren exposes the 24-7 lab culture familiar to many academic researchers. When she's desperate to get the papers and grants needed to keep open her fledgling lab at the Georgia Institute of Technology in Atlanta, Jahren and her technician and friend, Bill, go days at a time without eating or sleeping properly. At a seminar, Jahren gnaws on a dog biscuit to prevent her stomach from rumbling. In another account, a night janitor stumbles on her and wisely observes: "However much you love your job, it ain't gonna love you back." Meanwhile, the long-suffering Bill sleeps rough in a van because his meagre salary cannot stretch to renting a room.

Jahren also pulls no punches on the stark realities of being a woman in science, which won't come as a surprise to many. She defines sexism as "the cumulative weight of constantly being told that you can't possibly be what you are". At one conference, she stands alone, "trapped with pasty middle-aged men who regarded [her] as they would a mangy stray". She overhears male colleagues speculating about her sexual orientation and waistline, and has to endure colleagues' discomfort during her pregnancy. This is not a how-to manual, but young scientists of either gender could learn a lot simply from Jahren's perseverance.

Lab Girl is funny, full of joyous moments and often sad. But despite all the hardship, there is clearly nowhere else that Jahren would rather be. "My lab," she writes, "is a place where I can be the child I still am," a home where "the lights are always on," a church "because it is where I figure out what I believe". ■

Jennifer Rohn leads a cell-biology group at University College London. Her most recent novel is *The Honest Look*.
e-mail: jenny@lablit.com

Correspondence

Authenticate new xenograft models

As the US National Cancer Institute (NCI) switches to using patient-derived tumour xenografts in mice for drug screening (see *Nature* **530**, 391; 2016), we warn researchers on behalf of the International Cell Line Authentication Committee (go.nature.com/kphqtx) that xenografts potentially have the same cross-contamination and misidentification problems as cultured cell lines.

For example, at least 4 of the 60 human-cancer cell lines on NCI's panel are affected, as are several NCI-derived lines that were established before contamination testing became widely available (see go.nature.com/csodfc).

Lack of testing or bad practice can undermine the reliability of patient-derived xenografts. Cell cultures from these call for extra quality-control measures, such as screening for cross-species DNA contamination (J. Camps *et al.* *Leuk. Res.* **30**, 923–934; 2006).

The US National Institutes of Health now requires the authentication of key biological resources, but guidelines and defined protocols are still needed for rigorous characterization of patient-derived xenografts. It will be essential to deposit reference information for cell-based models in public databases, including details of donor DNA profile, host species and strain, testing outcome for host DNA, and methodologies.

Roland M. Nardone *Silver Spring, Maryland, USA.*

Roderick A. F. MacLeod *Leibniz Institute DSMZ, Germany.*

Amanda Capes-Davis *Children's Medical Research Institute, New South Wales, Australia.*
acapes-davis@cmri.org.au

Science primers in the courtroom

David Neuberger suggests that scientific primers could help to speed up legal proceedings

(*Nature* **531**, 9; 2016). In the United States, the Federal Judicial Center and the National Academies of Sciences, Engineering, and Medicine have prepared such primers for judges in the form of a reference manual (see go.nature.com/ntqkvf).

Judges in other countries have used this manual to understand disputes over scientific evidence. The Argentinian judiciary has prepared 15 similar guides modelled on chapters of the US reference manual. The Canadian judiciary, through its National Justice Institute, has used it to prepare four supplementary chapters that deal with specific concerns in Canadian courts. The judiciaries of China, Japan and several others are also exploring ways to incorporate sections of the US manual into judicial-education programmes.

As co-chairs of the National Academies' Committee on Science, Technology, and Law, we find that engagement between our communities creates new understandings and strengthens the foundation for decision-making within both groups. The two communities should seek more opportunities to discuss science in the courtroom and law in the laboratory.

David Baltimore *California Institute of Technology, Pasadena, USA.*

David S. Tatel *US Court of Appeals for the District of Columbia Circuit, Washington DC, USA.*
baltimo@caltech.edu

IPBES reaches out to social scientists

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) has issued a call for experts to perform its global assessment (see go.nature.com/g5zc3e). The call, which closes on 5 May, emphasizes the need to nominate more social scientists (see also A. B. M. Vadrot *et al.* *Nature* **530**, 160; 2016 and K. Reuter *et al.* *Nature* **531**, 173; 2016).

A strong collective effort is necessary to reach scholars outside the natural sciences, because they might not consider themselves to be biodiversity researchers. To this end, IPBES is reaching out to learned social-science societies (of sociology, economics, geography, anthropology, political science and psychology, among others), to networks of scientists and to prominent interdisciplinary international programmes such as Future Earth.

A special procedure to fill gaps in expertise, including in the social sciences, was adopted at the IPBES fourth plenary (further suggestions to secretariat@ipbes.net are welcome). Success ultimately depends on governments and organizations stepping up to nominate more social scientists.

Anne Larigauderie *IPBES Secretariat, Bonn, Germany.*

Marie Stenseke *University of Gothenburg, Sweden.*

Robert T. Watson *Potomac, Maryland, USA.*
secretariat@ipbes.net

Self-built labware stimulates creativity

Adapting open-source online designs for the assembly of scientific equipment is useful for cost-cutting (*Nature* **531**, 147–148; 2016). Just as important from an educational viewpoint is the resourcefulness and creativity that such 'open hardware' can stimulate in customizing it to address scientific goals.

My graduate students have used open-source electronic modules and 3D printing over the past three years to generate prototype devices for research in applied chemistry. Their limited experience in electronics is not a problem because the equipment is not expensive — plenty of open-source printed circuit boards have been discarded without denting the lab's budget. Crucially, the end products worked and furnished useful data (see also P. L. Urban

Analyst **140**, 963–975; 2015).

Pawel Urban *National Chiao Tung University, Hsinchu, Taiwan.*
plurban@nctu.edu.tw

A different take on Indus evidence

As a scholar of the Indus script since the 1990s, I take issue with many of the statements in Andrew Robinson's summary (*Nature* **526**, 499–501; 2015 and *Nature* **532**, 308; 2016) of his book *The Indus: Lost Civilizations* (Reaktion, 2015).

In my view, there are climatic, geological and inter-regional motivations for the end of urbanization in the Indus. I dispute that there is evidence for Hinduism's roots in the Indus Valley. And my inference from photos of the many axes, spearheads and arrowheads in site reports is that the Indus people had military weapons.

I consider standard usage of the term pictograph in archaeology to be a symbol, with no linguistic counterpart, representing a real or mythical object. A sign or character representing a word or phrase, such as those used in shorthand and some writing systems, is a logogram. My reading of the decipherments of Mayan and Linear B does not chime with Robinson's précis. Finally, I dislike the way he elides code-breaking and decipherment.

A point-by-point discussion of these and other issues can be found in my books (*Epigraphic Approaches To Indus Writing*; Oxbow, 2011; *The Archaeology and Epigraphy of Indus Writing*; Archaeopress, 2015), which demonstrate the potential of digital techniques to move this field forward.

The Indus peoples, like all other archaeological cultures, had the same range of foibles and brilliance we all share. We owe it to them to make our representations of their writing and culture as precise as possible.
Bryan K. Wells *Victoria, Canada.*
bdkj.wells@gmail.com

Geoffrey Eglinton

(1927–2016)

Pioneer of molecular–fossil research.

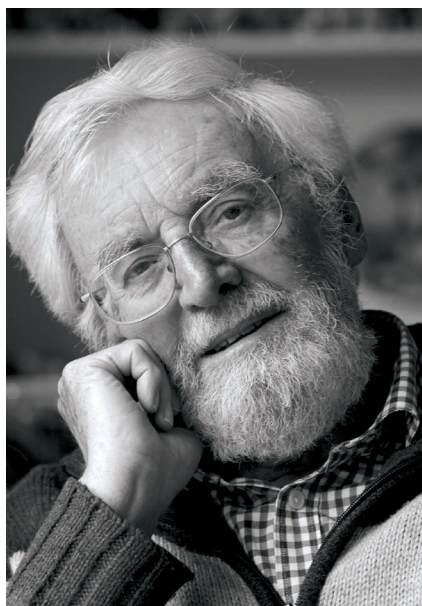
Geoffrey Eglinton was curious about the history of molecules. He followed their passage from living organisms into soils and sediments, and tracked their geological fate in sedimentary rocks and fossil fuels. His exploration of the natural history of biochemicals and their geochemical remnants established the modern field of organic geochemistry. In 1969, he analysed Moon rocks collected by Neil Armstrong and Buzz Aldrin on *Apollo 11*.

Eglinton, who died on 11 March, was born in Cardiff, UK, in 1927. He studied chemistry at the University of Manchester, from where he earned three degrees: a BSc in 1948, a PhD in 1951 and a DSc (a doctorate of science) in 1966. He worked for two years as a postdoctoral researcher at Ohio State University in Columbus and then returned to the United Kingdom as an Imperial Chemical Industries (ICI) fellow at the University of Liverpool. In 1954, he became a lecturer at the University of Glasgow.

Eglinton's original training was in synthetic chemistry. His early accomplishments included devising a new way to form carbon–carbon bonds by joining two compounds, each of which contained a carbon triple bond — a process now known as the Eglinton reaction.

His shift towards the chemistry of natural products, and ultimately to geochemistry, followed the arrival of a new analytical tool in the early 1950s: gas chromatography. The technique, which separates compounds carried by a gas along a liquid surface in a narrow column, proved invaluable to untangling complex mixtures of natural organic compounds. Eglinton was the first to use gas–chromatography separation in the analysis of chemicals called terpenoid lipids, which are found in plants as well as in ancient sediments. Soon, he became interested in the waxy lipids that cover the surfaces, or cuticles, of leaves, and began to determine their distributions.

Waxes protect leaves from water loss and from insects and fungi. During the late 1950s, Eglinton became fascinated with plant–wax compounds, which persist in soils, sediments, rocks and petroleum. In 1960, he took his young family to the University of La Laguna in Tenerife, Spain, for a sun-filled sabbatical. He wanted to discover whether different plant taxa have characteristic patterns of long-chain cuticular lipids; if they did, he knew that the compounds would be of enormous value in reconstructing the ecosystems of the past.



Eglinton's pioneering work elegantly wove together chemistry, biochemistry and botany, and culminated in a comprehensive paper published in 1967 in *Science* on leaf waxes, which is still a defining document in the field and Eglinton's most cited work (G. Eglinton and R. J. Hamilton *Science* **156**, 1322–1335; 1967). He studied the geochemistry of plant waxes for the rest of his career and well into his retirement. Indeed, his prescient admiration for plants' persistent waxes laid the foundation for their wide use today as palaeoclimate signatures.

In 1963, Eglinton began seeking molecules from the earliest life on Earth, in collaboration with the biochemist and Nobel laureate Melvin Calvin. Eglinton used his analytical expertise to search for biologically derived organic molecules in sedimentary rocks that were more than a billion years old. His work with Calvin revealed that early life had a biochemistry that was fundamentally similar to that of modern cells. The discovery of the startling antiquity of chemical remains from ancient cells sparked people's imaginations, and helped to introduce the concept of 'molecular fossils' to a broad audience.

During the mid-1960s, Eglinton's exquisite studies attracted the interest of researchers at NASA. They recognized that the ancient molecular fossils were definitive biosignatures and that organic geochemistry would be highly useful in studies of lunar samples.

Eglinton's team included the leading

organic geochemists of the day. The analytical detective work on the Moon rocks required extreme cleanliness to avoid contamination. So clean were the researchers' methods that they found minute traces of carbon from the solar wind blasted into lunar minerals. The work earned Eglinton the NASA Gold Medal for Exceptional Scientific Achievement and further elevated the growing field of organic geochemistry.

In 1967, Eglinton moved from Glasgow to the University of Bristol, where, with his friend and colleague James Maxwell, he established the Organic Geochemistry Unit (OGU). The OGU quickly became a global centre of excellence in organic geochemistry. Generations of students and postdocs studied fossil molecules there, which they used to study life in and trace the temperature of ancient oceans, and to probe oil transformation in geological basins.

Geoffrey, whom I knew professionally and through friendship with his family, always paid the highest compliment to young scientists: he listened intently to their ideas. After retiring from Bristol in 1993, he continued to work as an emeritus professor and through adjunct appointments with various institutions, including the Swiss Federal Institute of Technology in Zurich, where he often collaborated with his son, Timothy, a professor of biogeoscience and contemporary of mine.

Geoffrey published more than 500 papers and received numerous honours, including being elected fellow of the Royal Society in London. His greatest reward was the work itself and his many collaborations with those who shared his passion. His joy in the rich world of molecular fossils radiates from the pages of a 2008 book that he co-authored with Susan Gaines and Jurgen Rullkötter, *Echoes of Life* (Oxford University Press), which chronicles the science and the scientists that helped him to build the field of organic geochemistry.

Geoffrey was beloved by his wife of more than 60 years, Pam, his children, grandchildren and friends — and by his global scientific family working in the discipline that he founded. ■

Katherine H. Freeman is a distinguished professor of geosciences at Pennsylvania State University, University Park, Pennsylvania, USA. She knew and admired Geoffrey Eglinton for nearly 30 years, as a colleague and friend.
e-mail: khf4@psu.edu

ALICE CAVE

MICROBIOME

Eating for trillions

Three studies investigate the bacteria in the guts of malnourished children and find that, when this microbiota is transferred into mice, supplements of certain microbes or sugars from human breast milk can restore normal growth.

DERRICK M. CHU & KJERSTI M. AAGAARD

Childhood undernutrition accounts for nearly half of all deaths in infants under the age of five worldwide, and efforts to restore adequate nutritional intake in early infancy have produced only modest outcomes¹. A proposed alternative is to alter the population of bacteria that inhabits the intestine, known as the gut microbiota, which modulates intestinal metabolic activity. Indeed, childhood malnutrition has previously been associated with an altered microbiota². Three papers (two in *Science*^{3,4} and one in *Cell*⁵) now further establish that the composition of the microbiota is altered in chronically malnourished children. Furthermore, modifying these microbial communities in mice can ameliorate diet-associated growth deficits.

The time between conception and around three years of age is crucial for growth and development in humans. Decades of research into the developmental origins of health and disease has shown that environmental influences during this period can contribute to disease later in life. For instance, early-life nutritional deficits or excess can have lifelong consequences for metabolic and cardiovascular health⁶. Because of the gut microbiota's symbiotic role in intestinal metabolism, changes in its composition are thought to contribute to these health problems.

The food that we consume influences which intestinal microbes flourish. Each microbial species has an optimal metabolic milieu that preferentially supports its growth. As such, a diet rich in plant fibre promotes a gut microbiota considerably different from that promoted by a diet rich in animal fats⁷. In return, the microbiota converts otherwise indigestible dietary components, such as fibre, into useful compounds that fuel intestinal-cell growth and promote healthy immune development. But what happens to the gut microbiota when children are malnourished?

In the first of the three papers, Blanton *et al.*³ addressed this question. They studied the microbiota in stool samples collected from Malawian children with either no or varying degrees of growth impairment. They used the results from the children who grew normally to derive a model of healthy microbiota

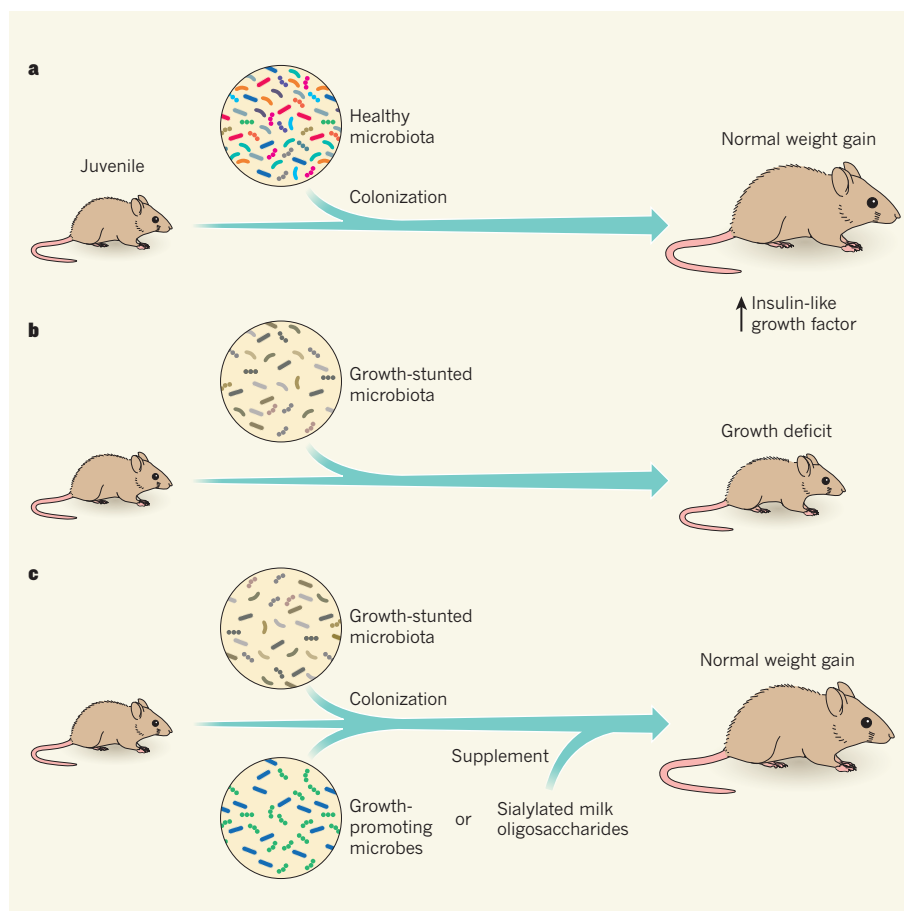


Figure 1 | Help for a healthy gut. Three studies^{3–5} analysed how changes in the bacteria that inhabit the gut (the gut microbiota) affect childhood growth. **a**, Juvenile mice whose guts are colonized by gut microbiota taken from healthy Malawian children experience robust weight gain, even if fed a nutrient-poor diet. This weight gain might be partially due to increased production of insulin-like growth factor protein, which is promoted by the presence of certain components of the microbiota. **b**, By contrast, mice harbouring microbiota from a child with impaired growth do not gain weight normally. **c**, By including two growth-promoting microbial species from the healthy microbiota in the colonization step, or by supplementing the animals' diet with sugars called sialylated milk oligosaccharides, healthy weight gain can be restored.

development. When they applied the model across the entire cohort, the authors found that children who had impaired weight gain tended to have an immature microbiota compared with those of a healthy weight.

Blanton and colleagues next investigated whether the difference in the gut microbiota contributed to growth impairment. They used bacteria isolated from the children's faecal samples to colonize the intestines of germ-free

mice, which are devoid of microorganisms. The authors then fed the mice a nutrient-poor diet reflecting the typical Malawian diet.

After several weeks, mice harbouring a microbiota derived from donors with impaired growth (which we will call growth-stunted mice) gained significantly less weight than the control group given the microbiota of healthy children. However, if the two groups were housed in the same cage, microbes from

the controls were rapidly transferred to the growth-stunted mice, restoring the animals' weight to healthy levels. The researchers traced this effect to two bacterial species — *Ruminococcus gnavus* and *Clostridium symbiosum*. When introduced into germ-free mice along with the microbiota from donors who had impaired growth, these species promoted a robust weight gain (Fig. 1).

To gain insight into how *R. gnavus* and *C. symbiosum* promoted growth, Blanton *et al.* examined the stools and livers of the mice using mass spectrometry. Levels of by-products of amino-acid breakdown were decreased in the livers of mice that had restored weight gain compared with those in growth-stunted mice. The authors speculate that the addition of the two species shifted the animals' metabolism away from energy extraction, which involves amino-acid breakdown, towards growth and lean-mass building.

In the second paper, Schwarzer *et al.*⁴ provide additional clues to how the gut microbiota might systemically stimulate growth in early childhood. The authors compared the growth of normal juvenile mice with that of mice raised germ-free, and found that the presence of the microbiota promoted growth by facilitating the host's production of insulin-like growth factor protein. If the mice were placed on a nutrient-poor diet, microbial stimulation of this growth factor partially ameliorated growth deficits. Not all strains of bacteria could promote growth in this manner — different strains of one species, *Lactobacillus plantarum*, had differing effects on growth. These data highlight the fact that each microbe interacts with its host differently, warning that the beneficial effects of bacteria found to promote growth in mice might not be translated to humans.

In the third paper, Charbonneau and colleagues⁵ provide evidence that cultivating specific strains of bacteria is not the only way to stimulate microbiota-dependent growth. The authors studied the breast milk of mothers who had a malnourished child, and found that their milk tended to have lower amounts of specialized sugars called milk oligosaccharides. These sugars are of interest because they are abundant in human milk, but not in cow's milk. To investigate whether the sugars promoted microbiota-dependent growth, the researchers generated growth-stunted mice using a similar strategy to that of Blanton and colleagues. Adding a specialized 'sialylated' milk oligosaccharide to the growth-stunted animals' diet increased weight gain considerably, whereas treatment with other sugar types, or treatment of mice with an uncolonized gut, did not.

Unlike the other studies, Charbonneau *et al.* could not attribute the beneficial effects of milk oligosaccharides to a specific microorganism. Addition of the sugars did not dramatically alter the abundance of any one organism, although it did increase transcription

of microbiotic genes involved in starch metabolism and other metabolic processes. The authors concluded that the resulting growth was likely to be a consequence of numerous complex interactions within the entire community. Additional studies will be required to delineate these interactions.

Supplementing an infant's diet with milk oligosaccharides might promote healthy growth in children who do not receive the sugars through their mother's milk. However, manufacture of these oligosaccharides has proved difficult, which may be a barrier to widespread use.

It is becoming increasingly apparent that our diet, gut microbiota and health are inextricably linked^{2,8,9}. We must be conscious that, when we make dietary interventions, we affect the growth of trillions of bacteria. These three papers are a marked leap forward in our understanding of how the gut microbiota contributes to growth impairment in malnourished children, but by no means have they found a magic bullet for tackling growth stunting. Diet is just one of the many factors that govern growth in early life¹⁰. Furthermore, because nutritional deficiencies can vary

between communities¹, further studies should be performed on a population-by-population basis. Such research could optimize feeding regimens to take into account the beneficial effects of the microbiota. ■

Derrick M. Chu and Kjersti Aagaard are in the Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, Texas 77030, USA. **K. A.** is also in the Departments of Molecular and Human Genetics, Molecular and Cell Biology, and Molecular Physiology and Biophysics, Baylor College of Medicine. e-mail: aagaardt@bcm.edu

1. Dewey, K. G. & Adu-Afaruwah, S. *Maternal Child Nutr.* **4**, 24–85 (2008).
2. Smith, M. I. *et al. Science* **339**, 548–554 (2013).
3. Blanton, L. V. *et al. Science* **351**, aad3311 (2016).
4. Schwarzer, M. *et al. Science* **351**, 854–857 (2016).
5. Charbonneau, M. R. *et al. Cell* **164**, 859–871 (2016).
6. Barker, D. J. *Br. Med. J.* **301**, 1111 (1990).
7. David, L. A. *et al. Nature* **505**, 559–563 (2013).
8. Ma, J. *et al. Nature Commun.* **5**, 3889 (2014).
9. Schulz, M. D. *et al. Nature* **514**, 508–512 (2014).
10. Dewey, K. G. *J. Nutr.* **131**, 1879–1880 (2001).

This article was published online on 13 April 2016.

CLIMATE SCIENCE

Misconceptions of global catastrophe

American attitudes to changing weather, and therefore to climate change, have been analysed on the basis of US migration patterns since the 1970s. The findings have implications for the success of global climate policies. [SEE LETTER P.357](#)

JOACIM ROCKLÖV

For decades, studies have described how climate change might have catastrophic global consequences. Yet motivating individuals to make climate-friendly decisions in daily life seems to be challenging, and insufficient action to mitigate climate change has been taken. On page 357 of this issue, Egan and Mullin¹ report a study that might help to explain why. From an analysis of domestic human-migration patterns in the United States, the authors conclude that the net effects of climate change up to now are perceived by Americans to have been positive. Unless future climate threats are better understood and communicated, this perception may limit further action — until it is too late.

Estimates of the effects of future climate change cannot fully encompass the complexity of the changes. As the economist Nicholas Stern put it², “It is these hard-to-predict impacts that are the most troubling potential

consequences of inaction”. This might partly explain why the risks of climate change have been insufficient to provoke the general public to become a low-carbon-emitting society. It has been reported³ that technological innovation, ‘green’ jobs, and an understanding of the societal impact and health benefits of climate-change mitigation can stimulate public change through personal engagement — although other studies (see ref. 4, for example) report that such simple reframing is unlikely to gain much support from the public. But whether attitudes towards changing weather can inspire greater public action has yet to be proved, even if those attitudes become increasingly negative in the future.

In their study, Egan and Mullin estimated the US public's weather preferences by using an index score that measures the extent to which US migration patterns are associated with weather at different locations, adjusting for confounding factors using a previously reported method⁵. The index thus



Figure 1 | Warm winter. Children in T-shirts walk past Christmas decorations in New York City in December 2015. Egan and Mullin¹ report that Americans regard warmer winters as a positive effect of climate change.

describes climate effects that are meaningful to people's daily lives, and which stimulate responses. It incorporates trade-offs between different weather conditions, taking into account seasonal variation that may influence migration.

More specifically, the index included information about January's maximum temperature, July's heat index (a measure that combines maximum temperature and relative humidity), July's mean relative humidity, annual precipitation and the number of days with precipitation per year⁵. The authors weighted their weather preference index (WPI) to better reflect the geographical density and distribution of populations across the United States. They did this using a method⁶ developed by geographers to determine the exposure of human populations to weather, in which the amount of exposure is scaled to the inverse of the distance of weather observations from population centres of US counties.

Egan and Mullin find that America's domestic-migration patterns reflect a general dislike for warm, humid summers but an appreciation of warm winters. This means that the effects of climate change since the 1970s are perceived overall to have been an improvement — potentially undermining public support for actions that limit climate change. Their findings were largely unaffected when data from different sets of studies were used to calculate the WPI. Notably, the authors also show that the warmer winters currently enjoyed by US populations (Fig. 1) probably foreshadow warmer (and therefore negatively perceived) summers in the future, and that future perceptions of the

weather will probably become even more negative in scenarios that involve higher greenhouse-gas emissions.

A strength of the study is that it isolates the effect of long-term changes in daily weather on behaviour. But it does not incorporate the effects of weather extremes, which might be seen as a limitation. Public opinion, and potentially behaviour, is affected by extreme events and may catalyse support for emissions reductions⁷ — although such effects can be short-lived, as the authors have shown previously⁸.

It matters globally if US populations feel that they have benefited from the effects of climate change up to now.

The period studied in the retrospective part of the analysis is relatively short by the standards of climate studies, encompassing only 40 years. Over such short periods, naturally occurring climate variability can confuse the climate-change signal. Similarly, the reported signal cannot be differentiated into the effects of environmental changes that cause local warming, such as urbanization, and the outcomes of global phenomena related to the greenhouse effect. The prospective part of the study is robust — the climate-change projection takes into account the predictions of 35 climate models.

Another limitation of the study is that populations are assumed to remain in counties for which the WPI indicates public displeasure with climate. According to the definition of the index, such displeasure should contribute

to migration out of those areas. The future, population-weighted change in the WPI is therefore likely to be less extreme than that reported, solely because people will move away from regions that develop unsatisfying climate conditions. On the other hand, the exaggerated population-weighted changes in the WPI might allow components of migration patterns to be estimated in future climate scenarios, assuming the weather preferences remain unchanged.

Egan and Mullin show that it may be hard for Americans to understand how changes in climate can be a catastrophe when such changes apparently make daily life more pleasant. The United States is one of the most influential countries in global policymaking, and one of the greatest emitters of greenhouse gases⁹. It therefore matters globally if US populations feel that they have benefited from the effects of climate change up to now. Moreover, many Americans may not recognize that these effects will change, as Egan and Mullin show, and that even if emissions are reduced now, it will be a long time — not until roughly the middle of the twenty-first century¹⁰ — before climate alters in response.

The study's findings might also be relevant to the attitudes of European populations. In colder regions, such as in Canada, Russia and China, future winter warming may also be perceived as a benefit of climate change, but the WPI would need to consider the effects of climate change on snowfall and ice formation, and how people respond to those changes. It would also be interesting to examine whether attitudes to climate change are different in tropical or subtropical regions of Asia and Africa — perhaps an initially positive experience has already reversed in some areas, causing displeasure or emigration from affected regions. More broadly, studies that provide insight into the public experience of weather will help inform global climate policy. ■

Joacim Rocklöv is in the Unit of Epidemiology and Global Health, Umeå Center for Global Health Research, Umeå University, Umeå 90187, Sweden. e-mail: joacim.rocklov@umu.se

1. Egan, P. J. & Mullin, M. *Nature* **532**, 357–360 (2016).
2. Stern, N. *Nature* **530**, 407–409 (2016).
3. Watts, N. et al. *Lancet* **386**, 1861–1914 (2015).
4. Bernauer, T. & McGrath, L. F. *Nature Clim. Change* <http://dx.doi.org/10.1038/nclimate2948> (2016).
5. Rappaport, J. *Reg. Sci. Urban Econ.* **37**, 375–398 (2007).
6. Hanigan, I., Hall, G. & Dear, K. B. G. *Int. J. Health Geogr.* **5**, 38 (2006).
7. Spence, A., Poortinga, W., Butler, C. & Pidgeon, N. F. *Nature Clim. Change* **1**, 46–49 (2011).
8. Egan, P. J. & Mullin, M. *J. Polit.* **74**, 796–809 (2012).
9. Matthews, H. D. et al. *Environ. Res. Lett.* **9**, 014010 (2014).
10. IPCC. *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2013).

EPIGENETICS

An elusive DNA base in mammals

The discovery of a modified version of the base adenine, known as N^6 -methyladenine, in mouse DNA puts paid to the theory that cytosine derivatives are the only modified bases in mammals. [SEE ARTICLE P.329](#)

GERD P. PFEIFER

The DNA of most organisms is composed of four standard bases and a small set of modified bases that are produced enzymatically from these four after DNA replication. One modified base, N^6 -methyladenine (N6mA), is prevalent in prokaryotes (bacteria and archaea), but whether it is found in mammals has remained unclear. In this issue, Wu *et al.*¹ (page 329) report the existence of N6mA in mouse stem cells. This exciting discovery is enhanced by the identification of an enzyme that removes methyl groups from N6mA, and by the finding that the modification is enriched in certain regulatory DNA sequences — data that together provide clues to N6mA's possible function in mammalian genomes.

The most abundant modified base in the mammalian genome is 5-methylcytosine (5mC), which regulates the expression of genes and of DNA regions called retrotransposons, which, through reverse transcription to produce an RNA intermediate, can move around the genome, disrupting gene regulation^{2,3}. Enzymes called TETs can oxidize 5mC to form 5-hydroxymethylcytosine and other derivatives⁴. In prokaryotic genomes, N^4 -methylcytosine and N6mA are also widespread, and are thought to modulate DNA replication and repair, among other roles⁵. Moreover, N6mA is found in certain unicellular eukaryotic (nucleus-bearing) organisms⁶.

It was long thought that N6mA was absent from the DNA of most multicellular organisms. However, this view was biased by a lack of sufficiently sensitive detection methods, which prevented in-depth analysis of rare modified bases. In 2015, three studies characterized eukaryotic N6mA in detail — in green algae⁷ (*Chlamydomonas reinhardtii*), in nematode worms⁸ (*Caenorhabditis elegans*) and in fruit flies⁹ (*Drosophila melanogaster*). These reports revealed that the prevalence of the modification varied widely between eukaryotic species, and found evidence that, in these three organisms, N6mA is associated with active transcription.

Wu *et al.* looked for N6mA in the DNA of mouse embryonic stem (ES) cells. Using mass spectrometry, they found that N6mA represented only 6–7 bases per million adenines genome-wide. N6mA was enriched

approximately fourfold in genomic regions associated with a rare histone protein, H2AX, although the reasons for this association remain unclear.

The authors next focused on DNA regions that are bound to this atypical histone protein, which is often associated with genomic regions involved in cell regulation and development. Analysing only these selected sequences enabled the researchers to make use of a sophisticated technique called single-molecule real-time (SMRT) sequencing, which detects the different kinetics with which a polymerase enzyme replicates modified bases compared with standard ones. SMRT sequencing has been used to map N6mA in several prokaryotic species¹⁰, and Wu *et al.* used the technique to identify consensus DNA sequences at which N6mA arises in mouse cells.

Demethylase enzymes remove methyl groups from DNA, RNA or proteins. A few N6mA demethylases of the ALKB protein family have been shown to remove methyl groups from N6mA in RNA — for example, to regulate messenger RNA¹¹. In mammals, this protein family has nine members¹². Wu *et al.* found that deletion of one member, ALKBH1, in mouse ES cells led to a high accumulation of N6mA in the genome, and that ALKBH1 could remove methyl groups from N6mA in DNA *in vitro*.

It is possible that other ALKBH proteins also possess DNA N6mA demethylase activity, because ALKBH4, 6 and 7 have no clearly defined substrate. Curiously, ALKBH1 worked most effectively on single-stranded DNA *in vitro*, raising the question of whether the enzyme preferentially operates during transcription or DNA replication *in vivo*, when DNA is transiently single stranded. Of note, the *C. elegans* N6mA demethylase NMAD-1, which is more closely related to ALKBH4 than to any other mammalian protein, does act on double-stranded DNA⁸. Thus, ALKBH4 might also be a mammalian N6mA demethylase, perhaps acting at a different stage of the cell cycle from ALKBH1.

Wu *et al.* deleted the *Alkbh1* gene in ES cells, and found that N6mA accumulated on the X chromosome in these cells. This accumulation occurred in the regulatory regions of evolutionarily young LINE-1 retrotransposon sequences (those that moved into position relatively recently), suggesting that N6mA may help to control these genomic parasites. Indeed, the presence of N6mA correlated with inhibited expression of both LINE-1 elements and their adjacent genes (Fig. 1). This finding raises the possibility that N6mA has a role in mammalian X inactivation, a phenomenon that compensates for the fact that females have two X chromosomes, compared with males' one.

The gene-silencing role of N6mA in mouse ES cells contrasts with its presumed role in activating gene transcription in simpler eukaryotes^{7–9}. A recent study independently detected N6mA in frogs and in mouse kidney DNA¹³, finding that the modified base was relatively depleted in protein-coding sequences. However, the effect of this depletion on transcription has yet to be determined.

Wu and colleagues' confirmation of N6mA in mammalian genomes raises many questions. For example, why does N^6 -adenine methylation occur in such seemingly different

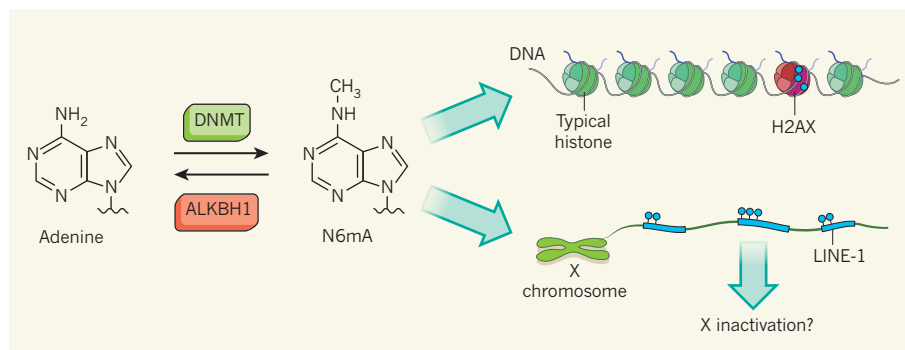


Figure 1 | Methyl modification in mammals. Methylation at the N^6 position of the DNA base adenine by a DNA methyltransferase (DNMT) enzyme produces the modified base N^6 -methyladenine (N6mA). Wu *et al.*¹ report that, in mammals, the demethylase enzyme ALKBH1 removes this methyl mark. The authors found that N6mA clustered in regions associated with the rare histone protein H2AX (N6mA denoted by blue circles) and in LINE-1 elements on the X chromosome. On the X, N6mA deposition correlated with reduced expression of LINE-1 elements and adjacent genes, indicating that the modification might be involved in controlling LINE-1 and in X-chromosome inactivation in mammals.

patterns and DNA sequences in different eukaryotic species? Is the modification differentially distributed across tissues or embryonic stages? An understanding of this variability might indicate whether there is a unifying role for N6mA in eukaryotes.

The enzymes that add methyl groups to adenine in mammalian DNA remain to be defined, as do the 'reader' proteins that detect genomic N6mA. Candidate reader proteins that can translate the N6mA code might be found among the several mammalian proteins that have domains predicted to interact with N6mA in DNA¹⁴.

ALKB-like proteins are currently the only known mammalian N6mA demethylases. In fruit flies, a relative of the mammalian TET 5mC oxidase acts as an N6mA demethylase⁹. However, the existence of 5mC in these

organisms is still controversial¹⁵, and the fruit-fly protein is structurally different from ALKB proteins. Thus, it remains unclear whether TET 5mC oxidase might act as an N6mA demethylase in mammals.

The levels of N6mA in multicellular organisms are exceptionally low. So, to begin to answer these many questions, the field must await further technological developments that facilitate the study of rare modifications. Such an ability would surely lift this blossoming research area to the next level. ■

Gerd P. Pfeifer is in the Center for Epigenetics, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA.

e-mail: gerd.pfeifer@vai.org

1. Wu, T. P. *et al.* *Nature* **532**, 329–333 (2016).
2. Bestor, T. H., Edwards, J. R. & Boulard, M. *Proc. Natl*

- Acad. Sci. USA* **112**, 6796–6799 (2015).
3. Schübeler, D. *Nature* **517**, 321–326 (2015).
4. Pastor, W. A., Aravind, L. & Rao, A. *Nature Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
5. Marinus, M. G. & Casadesu, J. *FEMS Microbiol. Rev.* **33**, 488–503 (2009).
6. Ratel, D., Ravanat, J. L., Berger, F. & Wion, D. *BioEssays* **28**, 309–315 (2006).
7. Fu, Y. *et al.* *Cell* **161**, 879–892 (2015).
8. Greer, E. L. *et al.* *Cell* **161**, 868–878 (2015).
9. Zhang, G. *et al.* *Cell* **161**, 893–906 (2015).
10. Murray, I. A. *et al.* *Nucleic Acids Res.* **40**, 11450–11462 (2012).
11. Zheng, G. *et al.* *Mol. Cell* **49**, 18–29 (2013).
12. Fedele, B. I., Singh, V., Delaney, J. C., Li, D. & Essigmann, J. M. *J. Biol. Chem.* **290**, 20734–20742 (2015).
13. Kozio, M. J. *et al.* *Nature Struct. Mol. Biol.* **23**, 24–30 (2016).
14. Iyer, L. M., Zhang, D. & Aravind, L. *BioEssays* **38**, 27–40 (2016).
15. Dunwell, T. L., McGuffin, L. J., Dunwell, J. M. & Pfeifer, G. P. *Cell Cycle* **12**, 3357–3365 (2013).

This article was published online on 30 March 2016.

20 years ago showed that the three transporters play an essential part in maintaining everyday control of neuronal signalling. Their significance has been further emphasized by the discovery⁴ that rare genetic mutations in the genes that encode them are linked to diseases such as autism, attention deficit hyperactivity disorder (ADHD) and parkinsonism.

Drugs such as cocaine, amphetamine and ecstasy exert their psychostimulatory action by hijacking these transporter proteins. Moreover, medications that target SERT, NET and DAT are used to treat depression and ADHD, among other conditions. Antidepressants, including selective serotonin reuptake inhibitors (SSRIs) and serotonin–noradrenaline reuptake inhibitors (SNRIs), block the transporters, thereby preventing reuptake of the neurotransmitter and consequently increasing its availability and so its overall activity in the synapse. In theory, this is how SSRIs and SNRIs alleviate some of the symptoms of depression

STRUCTURAL BIOLOGY

Antidepressants at work

Structures of the serotonin transporter protein SERT in complex with two different antidepressants shed light on how these drugs act, and point to possible targets for future drug development. [SEE ARTICLE P.334](#)

MARC G. CARON & ULRIC GETHER

On page 334 of this issue, Coleman *et al.*¹ describe the first high-resolution structure of the serotonin transporter (SERT) protein. This protein recaptures serotonin molecules that have been released by the cell and so modulates the effects of serotonin on neighbouring neurons. Those not familiar with SERT might question why solving this structure is such a big deal. First, it is technically challenging to purify the large amounts of SERT required for structural determination. Second, visualizing the detailed molecular structure of such a protein could provide unprecedented opportunities to develop more-selective and -efficacious therapies for diseases such as depression.

The neurotransmitter molecules serotonin, dopamine and noradrenaline are released from particular neurons and cross the synaptic cleft to bind to receptors on neighbouring neurons, where they modulate the effects of rapid excitatory or inhibitory signals from other neurotransmitters. SERT, together with the transporters for dopamine (called DAT) and noradrenaline (NET), belongs to the neurotransmitter:sodium symporter (NSS) family, which itself is part of the second-largest class of mammalian membrane-spanning protein — that is, proteins that move ions, small

molecules and nutrients into and out of cells. SERT, DAT and NET are selectively expressed on the surfaces of presynaptic neurons² and, by transporting serotonin, dopamine and noradrenaline back into the cell, act to terminate the modulatory effects of these neurotransmitters.

Genetic ablation studies³ conducted almost

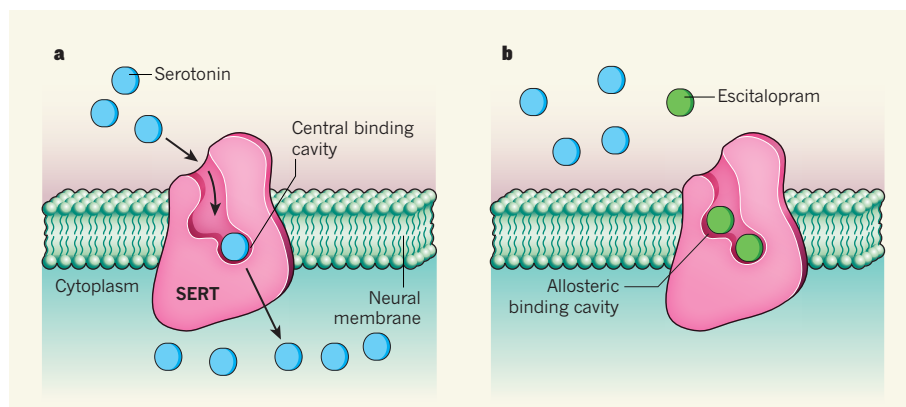


Figure 1 | Antidepressant blockade of a transporter protein. **a**, The neurotransmitter molecule serotonin is taken into neurons by the serotonin transporter (SERT) protein, through high-affinity binding in a central cavity. **b**, Coleman *et al.*¹ resolved structures of SERT bound to two antidepressants, either paroxetine or escitalopram (only escitalopram is shown). Both drugs bind in the central cavity, blocking serotonin transport by competing with the neurotransmitter for this binding site. Escitalopram also binds in a second, allosteric site, which is thought to prolong binding in the central site, enhancing the drug's SERT-blocking efficacy.

and other mood-related conditions.

In 2005, the group that carried out the current study determined the structure⁵ of LeuT, a bacterial relative of SERT. LeuT is involved in nutrient uptake in these unicellular organisms and functions similarly to SERT. In 2013, the same group also solved the structure of DAT in fruit flies⁶. This study suggested that the structural organization of these proteins is highly evolutionarily conserved, and defined several structural elements that indicated that the mechanism by which NSSs transport their substrates is also conserved. In providing molecular insights into the structure of human SERT, the current study confirms that nature, when it finds a way to do something, does it over and over again. But in the case of SERT, there is an interesting twist.

To solve structures at high resolution using X-ray crystallography, proteins must be purified in large quantities. Membrane-spanning proteins are notoriously difficult to purify, because they are unstable once removed from the hydrophobic-bilayer environment of the membrane. Previously, membrane proteins have been successfully stabilized by introducing mutations, by inducing the formation of a complex with easily crystallized proteins, or even by developing stabilizing antibodies to the native proteins⁷. Coleman *et al.* used an elaborate screen to painstakingly identify a few amino-acid residues that could be mutated to stabilize purified SERT without markedly affecting its functional properties. The resulting crystals revealed that the more than 600 amino-acid residues of SERT thread through the cell membrane 12 times to form an intricate 3D structure that is designed to mediate ion-coupled neurotransmitter transport across membranes.

The authors used X-ray crystallography to determine two structures of SERT bound to an SSRI — either paroxetine (Paxil) or escitalopram (Lexapro; the *S*-enantiomer of citalopram). As expected from the structure of fruit-fly DAT, they found that one molecule of paroxetine bound tightly in a central cavity in SERT that lies deep within the plasma membrane and is also thought to bind serotonin, thus competitively blocking serotonin transport. However, two molecules of escitalopram bound to SERT: one, like paroxetine, in the high-affinity central cavity, and the second more loosely in an outward-facing external vestibule (Fig. 1).

The existence of a second binding site (known as an allosteric site) away from SERT's primary binding site was proposed more than 30 years ago⁸. In 2012, modelling and mutational experiments⁹ suggested that the site was situated in the external vestibule, as Coleman *et al.* now confirm. The previous studies suggested that, when present at sufficiently high concentrations, escitalopram binding in the allosteric site could markedly retard the dissociation of drugs bound to the high-affinity site

— thus prolonging the SERT-blocking activity of escitalopram. This 'positive allosteric modifier' phenomenon has been proposed to explain the superior clinical efficacy of escitalopram compared with other SSRIs¹⁰, but confirming the mechanism in animal studies has been difficult¹¹. Coleman and colleagues' molecular documentation of the allosteric site is therefore important, because their structure might provide previously unappreciated opportunities for selective drug development.

The presence of allosteric sites on membrane proteins might be more common than anticipated. An example of the possible power of this finding comes from G-protein-coupled receptors (GPCRs) — membrane proteins that bind serotonin, dopamine, noradrenaline and many other signalling molecules on post-synaptic neurons. Successful determination of the structures of more than two dozen of these receptors has identified many potential allosteric binding sites that are being leveraged to help develop positive or negative modifiers of signalling to combat disease¹². The same opportunity could now present itself for transporters. Visualization of the molecular dynamics of transporter function, coupled with structure-based molecular modelling of drug binding, offers unprecedented opportunities for developing improved treatments for disorders of the central nervous system. ■

Marc G. Caron is in the Departments of Cell Biology, Medicine and Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA.

Ulrik Gether is in the Department of Neuroscience and Pharmacology, Panum Institute, University of Copenhagen, 2200 Copenhagen, Denmark.

e-mails: marc.caron@duke.edu; gether@sund.ku.dk

1. Coleman, J. A., Green, E. M. & Gouaux, E. *Nature* **532**, 334–339 (2016).
2. Kristensen, A. S. *et al. Pharmacol. Rev.* **63**, 585–640 (2011).
3. Torres, G. E., Gainetdinov, R. R. & Caron, M. G. *Nature Rev. Neurosci.* **4**, 13–25 (2003).
4. Ng, J., Papandreou, A., Heales, S. J. & Kurian, M. A. *Nature Rev. Neurol.* **11**, 567–584 (2015).
5. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. *Nature* **437**, 215–223 (2005).
6. Penmatsa, A., Wang, K. H. & Gouaux, E. *Nature* **503**, 85–90 (2013).
7. Kang, H. J., Lee, C. & Drew, D. *Int. J. Biochem. Cell Biol.* **45**, 636–644 (2013).
8. Plenge, P. & Møllerup, E. T. *Eur. J. Pharmacol.* **119**, 1–8 (1985).
9. Plenge, P. *et al. J. Biol. Chem.* **287**, 39316–39326 (2012).
10. Zhong, H., Haddjeri, N. & Sánchez, C. *Psychopharmacology* **219**, 1–13 (2012).
11. Jacobsen, J. P. *et al. Psychopharmacology* **231**, 4527–4540 (2014).
12. Conn, P. J., Lindsley, C. W., Meiler, J. & Niswender, C. M. *Nature Rev. Drug Discov.* **13**, 692–708 (2014).

This article was published online on 6 April 2016.

IMMUNOLOGY

Organelle stress triggers inflammation

The intracellular NOD1 and NOD2 receptors have been found to activate innate inflammation when a condition known as endoplasmic reticulum stress is induced by bacterial infection. SEE LETTER P.394

BENNETT H. PENN & JEFFERY S. COX

The ability of mammalian cells to recognize and respond to pathogenic invaders at the earliest stages of infection is crucial for immunity. Many host-cell receptors have been identified that recognize evolutionarily conserved microbial components, such as cell-wall peptidoglycan, to rapidly activate the innate immune system¹. In this issue, Keestra-Gounder *et al.*² (page 394) describe a curious connection whereby stress in the endoplasmic reticulum (ER), a membrane-bound compartment responsible for the synthesis of secreted cytokine proteins, triggers a cytoplasmic 'danger' signal leading to the release of inflammatory proteins that activate an innate immune response. Even more surprisingly, this ER

stress signal is transduced through NOD1 and NOD2 — sensors that respond to fragments of bacterial peptidoglycans — yet takes place in the absence of these molecules.

A wide variety of pathogens hijack and/or manipulate the ER of host cells³. Such infection often activates the stress-responsive control system of this organelle, termed the unfolded protein response (UPR)^{4,5}. Previous work indicates that activation of IRE1 α , a transmembrane kinase enzyme required for the UPR, leads to inflammation, although the mechanisms involved remain unclear^{6,7}. To avoid triggering the immune response, several bacterial pathogens actively inhibit the UPR³. By contrast, *Brucella abortus*, a bacterial pathogen that causes septic abortion in livestock, and which occasionally infects humans through contaminated dairy products, actively

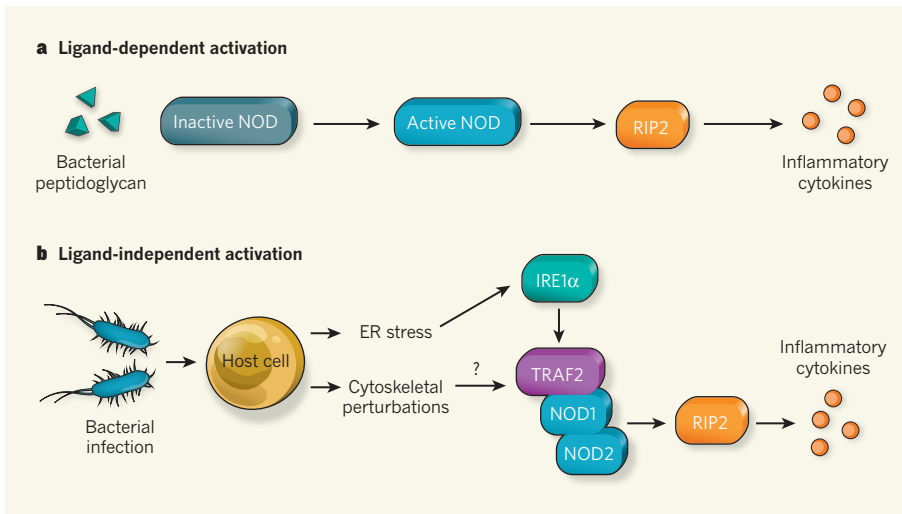


Figure 1 | NOD proteins respond to infection through distinct pathways. **a**, NOD1 and NOD2 were initially characterized as receptors that respond to bacterial peptidoglycan molecules in cells infected with bacteria, and that induce the production of inflammatory cytokine proteins through activation of the RIP2 kinase enzyme. **b**, Keestra-Gounder *et al.*² show that the cellular state known as endoplasmic reticulum (ER) stress, which is induced by various bacterial infections, generates a danger signal that is also transmitted by NOD proteins, even in the absence of bacterial products. The authors propose that ER stress stimulates the activity of the enzyme IRE1 α and leads to the formation of a complex between the enzyme TRAF2 and NOD1 and NOD2, initiating inflammation through RIP2 activation. When combined with previous observations that perturbations of the cytoskeleton, a common target of intracellular bacterial pathogens, may also signal through NODs to trigger inflammation^{15,17}, a model emerges in which NODs serve as a central hub for relaying signals that emanate directly from bacterial ligands and from stress pathways that indirectly detect pathogen attack on the cell.

promotes UPR-induced inflammation through its secreted effector protein VceC (ref. 8).

Keestra-Gounder *et al.* demonstrate that blocking either ER stress or the activity of the stress sensor IRE1 α during infection with *B. abortus* in mice markedly lowered production of the cytokine IL-6 and other mediators of inflammation. Conversely, triggering the UPR in mouse macrophage cells, either with chemical inducers or by expressing the *B. abortus* VceC protein, resulted in IL-6 release. However, a major twist came when the researchers investigated the mechanism by which the UPR triggered IL-6 release and identified the requirement for NOD1 and NOD2. Keestra-Gounder and colleagues found that genetic inactivation of both receptors, or deletion of the downstream enzyme RIP2, greatly reduced cytokine signalling in response to ER stress. Importantly, these sensors were activated in the absence of bacterial ligands, implying that NOD receptors can participate in both bacterial-ligand-dependent and -independent inflammatory responses (Fig. 1).

The authors then investigated UPR-mediated inflammation in *B. abortus* infection *in vivo*. When they treated pregnant mice infected with *B. abortus* with the bile salt TUDCA, a chemical used to mitigate the effects of ER stress, they observed dramatically damped inflammation in the placenta, and increased survival of newborn pups, without a change in the bacterial burden. *B. abortus* is usually transmitted between animals when a pregnant

female becomes infected, develops placental inflammation and has a spontaneous abortion, releasing the infected uterine contents. Animals grazing nearby subsequently eat these infected tissues and become infected. The finding that *B. abortus* secretes a protein that actively triggers placental inflammation suggests that it may be hijacking a host inflammatory pathway to promote its transmission to the next host.

The mechanism by which IRE1 α promotes inflammatory signalling remains unresolved. The authors propose a simple model in which IRE1 α stimulation leads to the formation of a complex between TRAF2 (a ubiquitin ligase enzyme that binds to IRE1 α), NOD1 and NOD2 (Fig. 1). This then leads to RIP2 activation, presumably by promoting interactions between these proteins. Although this is consistent with the requirement for TRAF2, NODs and RIP2 for inflammatory responses, the question of how the NOD receptors would be activated in the absence of bacterium-derived ligands remains untested. Although NODs have long been known to have a role in peptidoglycan sensing, only more recently have data emerged showing that peptidoglycan can directly bind NOD proteins⁹. It is possible that the peptidoglycan-independent signalling triggered by ER stress activates NODs through a distinct biochemical mechanism. An intriguing alternative possibility is that host-derived ligands are generated in response to ER stress, and activate NOD receptors in the cytoplasm.

Despite these uncertainties, Keestra-Gounder and colleagues' work compels researchers in the field to expand their thinking about the role of NODs in innate immunity beyond simply the sensing of bacterial ligands. It has long been recognized that the immune response to pathogens in plants relies in part on the monitoring of several normal host-cell physiological processes, and that disruption of these processes by a pathogen triggers an immune response¹⁰. It is now becoming increasingly evident that mammalian innate immunity also uses this strategy to 'guard' multiple cellular nodes commonly perturbed by pathogens; these include translation^{11,12}, membrane integrity¹³, mitochondrial function¹⁴, ER homeostasis⁵ and cytoskeletal integrity¹⁵. Strikingly, both the ER and cytoskeletal danger signals are relayed by the NODs in a ligand-independent manner, allowing NODs to participate in sensing a broad range of pathogens^{2,16}.

This ligand-independent NOD signalling also has potential implications for human health beyond the realm of infectious diseases. Several common chronic diseases, including type 2 diabetes and inflammatory bowel disease, are associated with ER stress and inflammation⁵, and might be therapeutically targeted through this newly identified ER–NOD pathway. ■

Bennett H. Penn is in the Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, San Francisco, California 94143, USA. **Jeffery S. Cox** is in the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720-3370, USA. e-mails: bennett.penn@ucsf.edu; jeff.cox@berkeley.edu

1. Kagan, J. C. & Barton, G. M. *Cold Spring Harb. Perspect. Biol.* **7**, a016253 (2015).
2. Keestra-Gounder, A. M. *et al. Nature* **532**, 394–397 (2016).
3. Janssens, S., Pulendran, B. & Lambrecht, B. N. *Nature Immunol.* **15**, 910–919 (2014).
4. Walter, P. & Ron, D. *Science* **334**, 1081–1086 (2011).
5. Garg, A. D. *et al. Trends Mol. Med.* **18**, 589–598 (2012).
6. Cho, J. A. *et al. Cell Host Microbe* **13**, 558–569 (2013).
7. Urano, F. *et al. Science* **287**, 664–666 (2000).
8. de Jong, M. F. *et al. mBio* **4**, e00418-12 (2013).
9. Grimes, C. L., Ariyandana, L. De Z., Melnyk, J. E. & O'Shea, E. K. *J. Am. Chem. Soc.* **134**, 13535–13537 (2012).
10. Jones, J. D. G. & Dangl, J. L. *Nature* **444**, 323–329 (2006).
11. Fontana, M. F. *et al. PLoS Pathog.* **7**, e1001289 (2011).
12. Vance, R. E., Isberg, R. R. & Portnoy, D. A. *Cell Host Microbe* **6**, 10–21 (2009).
13. Thurston, T. L. M., Wandel, M. P., von Muhlen, N., Foeglein, A. & Randow, F. *Nature* **482**, 414–418 (2012).
14. West, A. P. *et al. Nature* **520**, 553–557 (2015).
15. Kustermans, G. *et al. Biochem. Pharmacol.* **76**, 1214–1228 (2008).
16. Legrand-Poels, S. *et al. J. Cell Sci.* **120**, 1299–1310 (2007).
17. Keestra, A. M. *et al. Nature* **496**, 233–237 (2013).

This article was published online on 6 April 2016.

Distinct bone marrow blood vessels differentially regulate haematopoiesis

Tomer Itkin¹, Shiri Gur-Cohen¹, Joel A. Spencer^{2,3}, Amir Schajnovitz^{4,5,6}, Saravana K. Ramasamy⁷, Anjali P. Kusumbe⁷, Guy Lederger^{1,8}, Yookyung Jung^{2,3}, Idan Milo¹, Michael G. Poulos⁹, Alexander Kalinkovich¹, Aya Ludin¹, Orit Kollet¹, Guy Shakhari¹, Jason M. Butler⁹, Shahin Rafii⁹, Ralf H. Adams⁷, David T. Scadden^{4,5,6}, Charles P. Lin^{2,3} & Tsvee Lapidot¹

Bone marrow endothelial cells (BMECs) form a network of blood vessels that regulate both leukocyte trafficking and haematopoietic stem and progenitor cell (HSPC) maintenance. However, it is not clear how BMECs balance these dual roles, and whether these events occur at the same vascular site. We found that mammalian bone marrow stem cell maintenance and leukocyte trafficking are regulated by distinct blood vessel types with different permeability properties. Less permeable arterial blood vessels maintain haematopoietic stem cells in a low reactive oxygen species (ROS) state, whereas the more permeable sinusoids promote HSPC activation and are the exclusive site for immature and mature leukocyte trafficking to and from the bone marrow. A functional consequence of high permeability of blood vessels is that exposure to blood plasma increases bone marrow HSPC ROS levels, augmenting their migration and differentiation, while compromising their long-term repopulation and survival. These findings may have relevance for clinical haematopoietic stem cell transplantation and mobilization protocols.

Vascular-forming endothelial cells form a vast network which participates in homeostasis and metabolism regulation, delivering oxygen, nutrients and other building blocks to distinct organs. This diverse network also serves as a cellular highway allowing trafficking of blood cells, leukocytes and other cell types throughout the body. In addition, endothelial cells serve an important role as regulators of organ homeostasis and regeneration via direct interactions with local stem and progenitor cells, and by secretion of angiocrine factors¹. BMECs form a mechanical barrier, which prevents entry into the bone marrow of mature red blood cells and platelets from the circulation, regulating cellular trafficking, haematopoiesis and osteogenesis^{2–4}. BMECs also contribute to specialized perivascular microenvironments where the majority of bone marrow haematopoietic stem and progenitor cells (HSPCs) reside^{5–8}. BMEC perivascular domains include heterogeneous populations of mesenchymal stromal precursor cells (MSPCs) previously reported to regulate HSPCs^{9–11}. In addition, BMECs provide angiocrine signals that regulate haematopoietic stem cell development and haematopoiesis^{10,12,13}. Different types of blood vessels compose the bone marrow vascular network^{4,11,12}, exhibiting distinct properties and forming unique domains. We investigated how BMECs exert their dual roles as regulators of stem cell maintenance and of cellular trafficking, and if these distinct roles are associated with specialized blood vessels sub-types and specific micro-anatomical regions. We characterized the bone marrow vascular architecture, the properties of the distinct types of blood vessels, and their associated ‘niche’ cells participating in the formation of unique bone marrow multi-cellular domains. Finally, whether manipulation of endothelial properties may serve to control tissue homeostasis and stem cell fate was examined.

Defining bone marrow vascular architecture and domains

We used Sca-1-EGFP (Sca-1, also known as Ly6a) transgenic mice to distinguish between Sca-1⁺ sinusoidal BMECs (sBMECs) from Sca-1⁺

arterial BMECs (aBMECs)¹². Arterial BMECs (23.5 ± 3.1% of BMECs, Fig. 1a) display unique elongated elliptical nuclear morphology (Fig. 1b, f). Adherence and tight junction molecules VE-cadherin and ZO-1 were highly and preferentially expressed by aBMECs (Fig. 1c and Extended Data Fig. 1a). Sca-1⁺ blood vessels had smaller diameters compared to neighbouring Sca-1⁺ sinusoids, and were closely associated with calcified bone at the metaphysis or in the diaphysis (Fig. 1d and Supplementary Video 1). Arteries co-stained for Sca-1 and CD31, and were enwrapped by α SMA⁺ pericytes (Fig. 1e). Approaching the endosteum, arteries branched into smaller arterioles that were not associated with α SMA⁺ pericytes, but were instead surrounded by Sca-1⁺ mesenchymal (reticular) and clusters of Sca-1⁺ haematopoietic (round) cells (Fig. 1e). Combining osteopontin (OPN) staining for bone-lining osteoblasts with endothelial markers (Extended Data Fig. 1b), revealed that the vast majority of arterial blood vessels are found at a distance of <40 μ m from the endosteum, with ~50% at a closer distance of <20 μ m from the endosteum (Extended Data Fig. 1c). Arteries enwrapped by α SMA⁺ pericytes had an approximately 10 μ m diameter, branching to smaller ~5 μ m diameter endosteal arterioles, connecting downstream to much larger ~25 μ m sinusoids (Extended Data Fig. 1d).

High intracellular levels of reactive oxygen species (ROS) hamper bone marrow haematopoietic stem cell quiescence, accelerating their differentiation and exhaustion^{14,15}. Elevated ROS levels also promote HSPC mobilization by activating their motility machinery^{16,17}. Detecting intracellular ROS^{high} expressing cells (Extended Data Fig. 1e, f), we found there were lower frequencies of ROS^{high} cells adjacent to arteries (Fig. 1f (top and middle) and Extended Data Fig. 1g), compared to sinusoids (Fig. 1f (bottom) and Extended Data Fig. 1g). SLAM HSPCs associated with sinusoids and positioned >20 μ m from aBMECs (41 out of 41), had either negative or low (26 out of 41) or high (15 out of 41) ROS levels (Extended Data Fig. 1h, i). In contrast, SLAM HSPCs associated with arteries were consistently

¹Department of Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel. ²Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. ³Center for Systems Biology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁴Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Harvard Stem Cell Institute, Cambridge, Massachusetts 02114, USA. ⁶Center for Regenerative Medicine and Cancer Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁷Max Planck Institute for Molecular Biomedicine, Department of Tissue Morphogenesis and Faculty of Medicine, University of Münster, D-48149 Münster, Germany. ⁸Internal Medicine Department, Tel-Aviv Sourasky Medical Center, Tel-Aviv 64239, Israel. ⁹Department of Genetic Medicine, Weill Cornell Medical College, New York, New York 10065, USA.

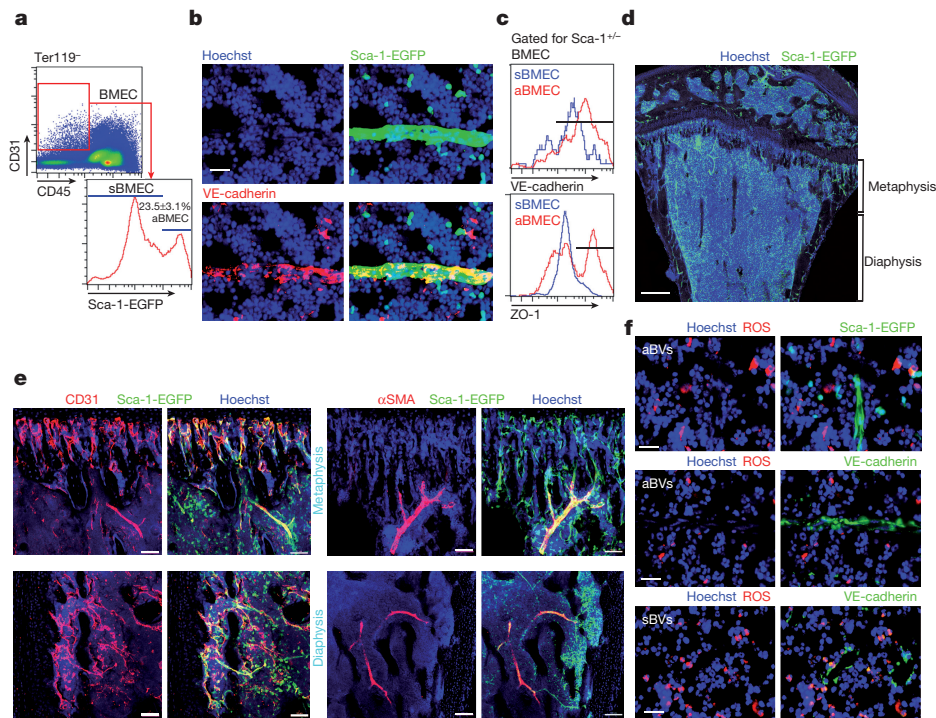


Figure 1 | Sca-1 and nestin distinguish less permeable arterial bone marrow blood vessels, which sustain ROS^{low} HSC. **a**, Representative flow cytometry density and histogram plots for BMECs (mean \pm s.e.m., $n = 6$ mice from three independent experiments). **b**, Representative fluorescence images of a small diameter blood vessel from the metaphyseal area expressing Sca-1-EGFP (green), junctional VE-cadherin (red) and elongated nuclei (Hoechst, blue). Scale bar indicates 20 μ m. **c**, VE-cadherin and ZO-1 flow cytometry representative histogram plots for mean fluorescent intensity expression (MFI) by BMECs. ($n = 9$ mice from three independent experiments). **d**, Representative confocal tile scan of

Sca-1-EGFP (green) femur. Scale bar indicates 300 μ m. **e**, Representative confocal images of endosteal regions in the metaphysis and diaphysis showing Sca-1⁺ (green)/CD31⁺ (red) arterial blood vessels and α SMA⁺ (red) pericytes. Scale bars in panels indicate 50 μ m, except for the lower right panels where scale bar indicates 100 μ m. **f**, Representative fluorescence images of ROS^{high} (red) expressing cells in the proximate microenvironment of Sca-1⁺ (green) or VE-cadherin⁺ (green) arterioles (top and middle panels, respectively) and VE-cadherin⁺ (green) sinusoids (bottom panels). Scale bar indicates 20 μ m.

negative for ROS staining (22 out of 22) (Extended Data Fig. 1j). The ROS^{low} SLAM HSPCs distant from arteries, resided in close proximity to megakaryocytes (<20 μ m) consistent with the quiescent megakaryocytic stem cell niche reports^{18–20} (Extended Data Fig. 1h). Our data indicate that HSPCs spatial bone marrow localization is dependent on HSPC metabolic state. However, in agreement with a recent study²¹, we confirmed that, if their metabolic state is ignored, SLAM HSPCs had random localization among distinct bone marrow regions (Extended Data Fig. 1k).

BMECs and neighbouring MSPCs characterization

As specific sub-types of endosteal blood vessels serve as regulatory niches for bone-forming mesenchymal stromal precursor cells (MSPCs)⁴, we analysed blood vessels for expression of the pericyte and MSPC marker NG2 (also known as CSPG4 in humans) and nestin^{9,11,13}. NG2 stained a heterogeneous population, negative and positive for Sca-1, enwrapping Sca-1⁺ blood vessels (Extended Data Fig. 2a). Unexpectedly, Sca-1 staining on bone marrow sections from mice with a nestin promoter driving GFP expression (nestin-GFP), reported to label HSC-supportive bone marrow MSPCs⁹, revealed the existence of nestin⁺ BMECs with elongated nuclei expressing Sca-1 and VE-cadherin (Extended Data Fig. 2b, c). Nestin⁺ precursor cells give rise to both endothelial and mesenchymal lineages²², and we have occasionally detected very bright nestin⁺Sca-1[−] MSPCs¹¹ adjacent to nestin⁺Sca-1⁺ BMECs (Extended Data Fig. 2b). Nestin⁺NG2⁺ MSPCs were tightly associated with nestin⁺NG2[−] blood vessels (Extended Data Fig. 2d). Nestin⁺ blood vessels are mostly small diameter, di-acetyl low-density lipoprotein (LDL)-negative blood vessels, associated with calcified bone (Extended Data Fig. 2e and Supplementary Video 2). Among CD45[−]CD31⁺ BMECs, nestin

expression was restricted to Sca-1⁺CD31^{high} aBMECs, and was absent from Sca-1[−]CD31⁺ sBMECs (Extended Data Fig. 2f, g). Nestin-GFP was also expressed by heterogeneous populations of mesenchymal⁹ and haematopoietic²³ populations (Extended Data Fig. 2g). Nestin⁺ blood vessels and MSPCs were predominantly detected at the metaphysis and adjacent to cortical bone at the diaphysis (Extended Data Fig. 2h, i). Nestin expressing non-myelinating Schwann cells that maintain hibernating HSC²⁴, were exclusively associated only with arteries, contributing to metabolically 'low' microenvironments (Extended Data Fig. 2j–l).

Further profiling expression of endothelial molecules involved in leukocyte trafficking, revealed that aBMECs express higher levels of VCAM-1, ICAM-1, P-selectin and preferentially expressed JAM-A, whereas sBMECs preferentially express higher levels of E-selectin (Extended Data Fig. 3a–f), which is involved in HSPCs bone marrow homing⁶. In addition to their role in leukocyte trafficking, these adhesion molecules also control HSPC retention by VCAM-1 (ref. 25) or negative regulation of HSC quiescence by E-selectin²⁶. Endothelial metabolism, more specifically glycolysis rather than oxidative phosphorylation, regulates vessel development and function^{27,28}. We noted that aBMEC exhibit lower ROS levels and higher glucose uptake relative to sBMEC (Extended Data Fig. 3g, h). This finding may indicate that aBMECs represent a more actively developing sub-type of BMEC and correlates with the observation that bone marrow oxygen tension is highest near and in arterial BVs²⁹ skewing arteries to the glycolytic pathway to avoid excessive ROS production²⁸.

Sinusoids are the exclusive trafficking sites

Relatively to sinusoids, the preceding arterial nestin⁺ blood vessels displayed lower permeability and dramatically higher blood flow and

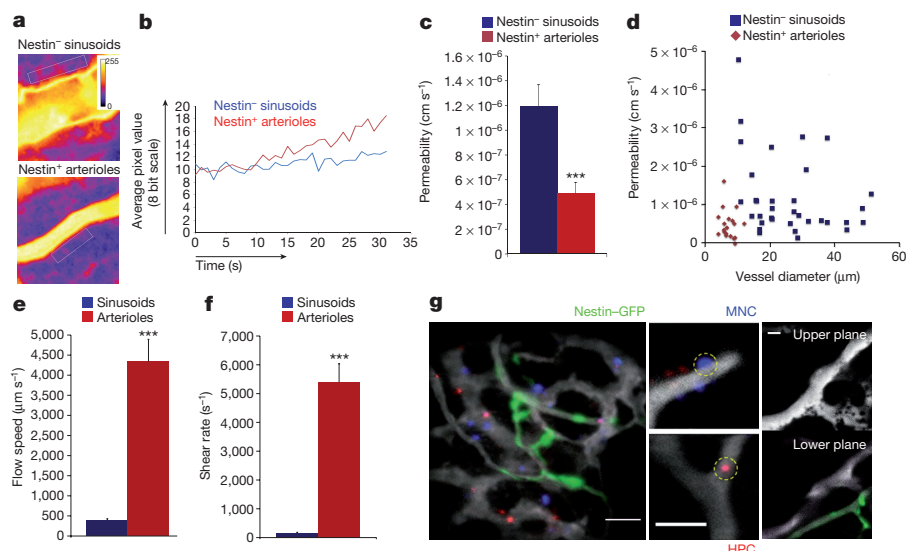


Figure 2 | Leaky sinusoids are the exclusive site for cellular trafficking. **a**, Standard deviation heat map of rhodamine-dextran leakage. Colour scale shows pixel intensity over 30 s of data acquisition. Regions of interest are bordered with a white line. **b**, Time traces of fluorescence signal. **c**, Average vascular permeability. **d**, A plot of permeability measurements as a function of vessel diameter. **e**, **f**, Average blood flow speed and shear rate. **g**, Representative maximum intensity images demonstrating homing of adoptively transferred lineage depleted HPC (haematopoietic progenitor cells, red) and bone marrow mononuclear cells (MNC) (blue) within the bone marrow of a nestin-GFP (green) mouse (left);

transmigrating bone marrow MNC (circled, blue) and adherent Lin[−] HPC (circled, red), scale bar indicates 50 μ m (middle); upper and lower planes of the z-stack taken from Supplementary Video 9 representing the relative proximity of a nestin⁺ blood vessel to a sinusoidal trafficking site (right), scale bar indicates 25 μ m, $n = 309$ trafficking events were monitored and analysed. **a–g**, Mean \pm s.e.m., $n = 53$ blood vessels were analysed for permeability measurements and $n = 62$ blood vessels were analysed for blood flow/shear rates, from three independent experiments each). Two-tailed Student's *t*-test; *** $P < 0.005$.

shear rates (Fig. 2a–f and Supplementary Videos 3–5). Comparing calvarial to femoral (including both diaphysis and metaphysis regions) bones, no major differences were noted in blood vessel distribution, blood vessel size, sinusoidal blood vessel properties or permeability, but the calvarial bone contained a higher frequency of BMECs, including a higher frequency of aBMEC, displaying characteristics of enhanced integrity (Extended Data Fig. 4a–e).

Phenotypically defined long-term repopulating (LTR)-HSCs are reported to localize preferentially near nestin⁺ stromal cells⁹. The lower permeability, the high-flow speed²⁹, and high shear rates in arterial nestin⁺ blood vessels suggest that a different type of blood vessel serves as a site for HSPCs and mature leukocyte trafficking and homing to the bone marrow. We observed that haematopoietic cell rolling and adhesion events occurred exclusively in sinusoids (377 out of 377

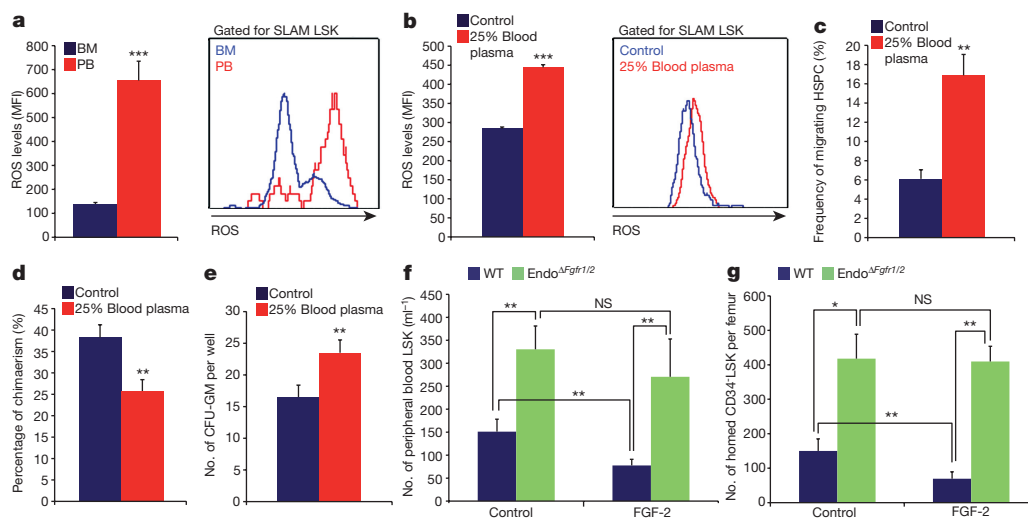


Figure 3 | Plasma penetration through leaky endothelium dictates HSPC trafficking and development. **a**, HSPCs ROS MFI quantitative analysis and representative histogram plot. Mean \pm s.e.m., $n = 6$ mice from three independent experiments. Two-tailed Student's *t*-test, *** $P < 0.005$. **b–e**, Bone marrow cells were incubated for 2 h with (25% blood plasma) or without (control) peripheral blood plasma. Mean \pm s.e.m., $n = 9$ repeats in triplicates from three independent experiments. Two-tailed Student's *t*-test; ** $P < 0.01$, *** $P < 0.005$. **b**, HSPCs ROS MFI quantitative analysis and representative histogram plot. **c**, HSPC migration frequency.

d, Chimaerism levels indicating LTR-HSC contribution. Mean \pm s.e.m., $n = 27$ donor from three independent experiments, with at least 3 recipient mice per donor. **e**, Average number of colony-forming units/granulocytes/macrophages (CFU-GM). **f**, Numbers of LSK HSPCs in the blood. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-way ANOVA with Bonferroni's multiple comparison post-hoc test; ** $P < 0.01$. **g**, HSPC homing per femur. Mean \pm s.e.m., $n = 5$ mice from 2 independent experiments. Two-way ANOVA with Bonferroni's multiple comparison post-hoc test; * $P < 0.05$, ** $P < 0.01$.

events, Supplementary Video 6 and Extended Data Fig. 4f, g). Similarly, transendothelial migration of mature leukocytes and immature HSPCs occurred exclusively via sinusoids (observed in 309 out of 309 events, Fig. 2g and Supplementary Videos 7–10).

Endothelial CXCL12 and CXCR4 in HSC mobilization

AMD3100 treatment, an agent that induces rapid HSPC mobilization³⁰ via shedding and release of chemoattractant cytokines, such as stem cell factor (SCF) and CXCL12, from bone marrow cells into the blood^{31,32}, reduced vascular integrity and cytokine abundance, preferentially on sBMECs (Extended Data Fig. 5a–d). Furthermore, AMD3100 led to reduced CXCR4 phosphorylation by BMEC, 5 min after treatment (Extended Data Fig. 5e). CXCR4 neutralization treatment increased vascular permeability as well (Extended Data Fig. 5f). However, since these antibodies also directly inhibit HSPC egress and mobilization³², *Cxcr4* was specifically deleted in endothelial cells (Endo $\Delta Cxcr4$, Extended Data Fig. 5g–j). Increased vascular permeability and enhanced HSPC egress were observed in Endo $\Delta Cxcr4$ mice (Extended Data Fig. 5k–m). Supported by previous *in vitro* findings³³, our data indicate that CXCR4 signalling regulates bone marrow vascular integrity and, as a consequence, HSPC trafficking.

A leaky endothelium promotes HSPC trafficking

To delineate a potential connection between blood vessel permeability allowing penetration of blood plasma into the bone marrow and

HSC development, bone-marrow-residing SLAM LSK HSPCs were compared with circulating SLAM LSK HSPCs in the peripheral blood. ROS levels in peripheral blood circulating HSPCs were much higher relative to bone-marrow-residing HSPCs (Fig. 3a). Short *in vitro* exposure of bone marrow HSPCs to peripheral blood plasma augmented intracellular ROS levels, resulting in their enhanced migratory capacity (Fig. 3b, c). Yet it also enhanced the frequency of apoptotic HSPCs, slightly increased HSPC cycling, along with enhanced differentiation, leading to reduced LTR-HSC capacity (Fig. 3d, e and Extended Data Fig. 6a, b).

We tested whether the state of endothelial integrity affects bone marrow HSC fate. Fibroblast growth factor (FGF) signalling has important roles in long-term repopulating HSCs (LTR-HSCs) maintenance and expansion^{34,35} and in maintaining endothelial integrity^{36,37}. Consistent with this, induction of FGF signalling enhanced bone marrow endothelial barrier integrity (Extended Data Fig. 6c–e). In addition, it led to significant changes in bone marrow vascular architecture (Extended Data Fig. 6f–j). As a consequence, HSPCs and MSPCs fates were affected, resulting in their expansion³⁴, reduced HSPCs and LTR-HSCs bi-directional trafficking, reduced MSPC differentiation and a shift in HSPC metabolism (Extended Data Fig. 6k–q).

To segregate endothelial mediated effects, FGFR1 and FGFR2 were deleted specifically in endothelial cells (Endo $\Delta Fgfr1/2$) (Extended Data Fig. 5g). Impairment in endothelial integrity (Extended Data Fig. 7a–c) was measured in Endo $\Delta Fgfr1/2$ mice. Consistent with their impaired

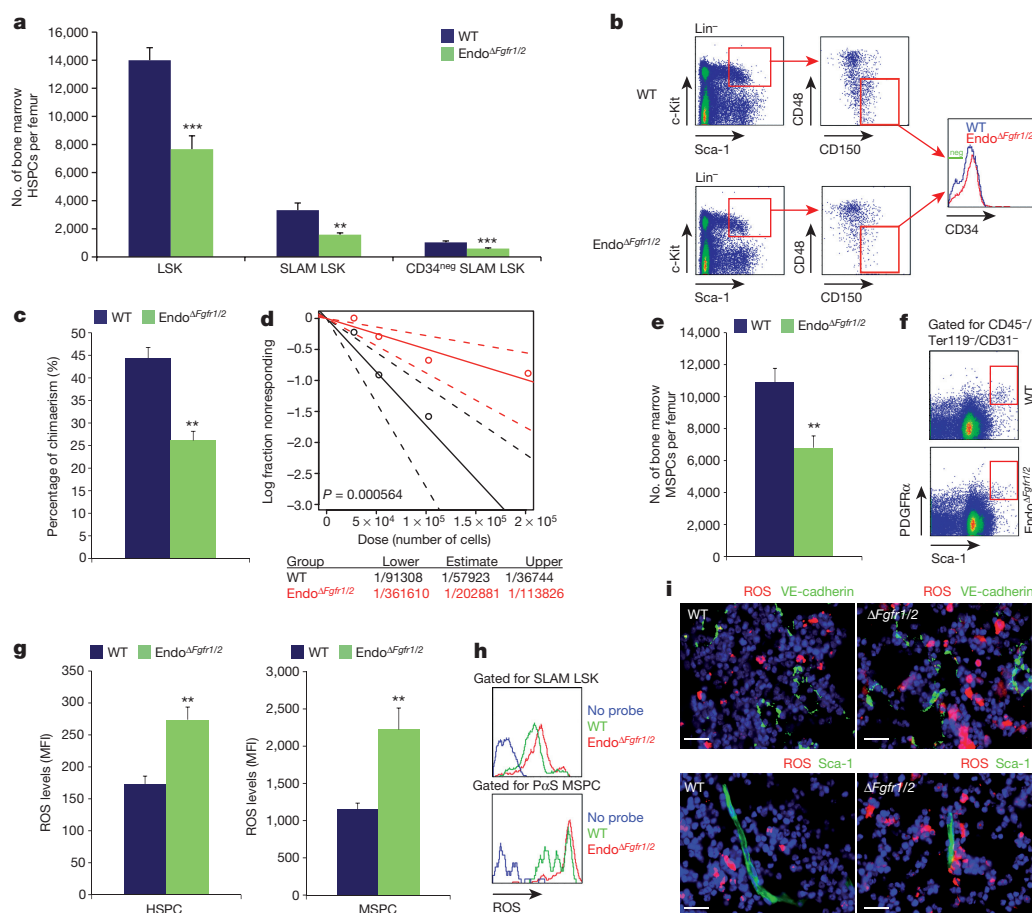


Figure 4 | Reducing endothelial barrier integrity hampers stem cell maintenance. **a, b**, Quantitative analysis and representative plots of HSPC populations. **c**, Chimaerism levels indicating LTR-HSC contribution. Mean \pm s.e.m., $n = 18$ donor mice from two independent experiments, with 3 recipient mice per donor. **d**, Repopulating units frequency in the bone marrow indicating HSC numbers ($n = 60$ donor mice from 2 independent experiments, for 3 recipient mice per dilution per donor).

e, f, Quantitative analysis and representative density plots of MSPCs. **g, h**, Quantitative analysis and representative histogram plots of HSPCs and PoS MSPCs ROS MFIs. **i**, Representative images of ROS^{high} (red) cells in proximity to blood vessels. Scale bar indicates 20 μ m. **a–h**, Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's *t*-test; ** $P < 0.01$, *** $P < 0.005$.

barrier status, $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice demonstrated increased HSPC bi-directional trafficking (Fig. 3f, g). In contrast to wild-type (WT) mice, FGF-2 treatment of $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice failed to reduce or prevent this phenomenon (Fig. 3f, g). Also, this model of endothelial FGF loss-of-function displayed notable changes in bone marrow vascular architecture and properties (Extended Data Fig. 7d–j).

Vascular integrity maintains bone marrow stem cells

We next investigated bone marrow HSC maintenance under conditions of impaired endothelial integrity. Endothelial barrier disruption in $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice led to a significant reduction in the numbers of HSPCs/LTR-HSCs and bone marrow $\text{PDFGR}\alpha^+\text{Sca-1}^+$ ($\text{P}\alpha\text{S}$) MSPCs³⁸ (Fig. 4a–f). Both HSPCs and MSPCs displayed increased ROS levels (Fig. 4g, h), and the frequency of ROS^{high} cells surrounding blood vessels was increased (Fig. 4i and Extended Data Fig. 7k).

Along with elevation in ROS levels, glucose uptake was enhanced by both HSPCs and MSPCs (Extended Data Fig. 7m and 8a) in $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice, indicating an augmentation of the oxidative phosphorylation pathway in these populations. Similar to plasma-exposed bone marrow HSPCs, HSPCs derived from $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice bone marrow exhibited a slight elevation in cycling state, a striking increase in apoptosis, and following transplantation, a differentiation skewing towards the myeloid lineage (Extended Data Fig. 7n–q). The functional role of elevated ROS levels in HSPCs from $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice was tested using the ROS scavenger *N*-acetyl-L-cysteine (NAC). Prolonged NAC treatment restored normal levels of HSPC egress to the peripheral blood, and normal levels of bone marrow HSPC and LTR-HSC in $\text{Endo}^{\Delta\text{Fgfr}1/2}$ mice (Extended Data Fig. 7r–t).

As MSPC metabolism and glucose uptake regulate bone formation³⁹, and plasma-borne factors such as vitamin E can penetrate the bone marrow and influence bone remodelling⁴⁰, stromal development was examined under conditions of a hampered endothelial bone marrow barrier. Accelerated stromal differentiation at the expense of the MSPC pool was observed (Extended Data Fig. 8a–d) alongside with dramatic changes in the bone marrow concentrations of the hormones calcitonin and parathyroid hormone (PTH), regulating bone formation and remodelling, responding to changes in endothelial integrity (Extended Data Fig. 8j–m).

We also applied a pharmacological model to disrupt endothelial integrity by infusing neutralizing anti-VE-cadherin antibodies (Extended Data Fig. 9a). This model mimicked our genetic model affecting HSPC bi-directional trafficking and LTR-HSC maintenance which were also ROS dependent (Extended Data Fig. 9b–k). Vascular architecture and stromal development were also severely altered in this model of barrier manipulation (Extended Data Figs 7l, 8e–i, and 9l–q).

Discussion

In this study, the roles of BMECs as regulators of haematopoiesis were investigated. The existence of diverse vascular niches was investigated and an endosteal-vascular niche is supported by data showing that less permeable endosteal blood vessels provide a microenvironment promoting ROS^{low} HSC maintenance (summarized in Extended Data Fig. 10). In support of this, specialized aPC-secreting bone marrow arteries retain EPCR^+ LTR-HSC via downregulation of nitric oxide production, enhancing adhesion and reducing migration⁴¹. These data sets add new information to previous reports that BrdU -retaining HSCs are mostly localized in endosteal regions⁴², where HSCs are maintained in a quiescent mode^{43,44}. We extended previous studies reporting similar niches at steady-state²¹ and in irradiated transplanted recipient mice^{8,45}, located mainly at trabecular regions between the endosteal surface and previously undefined blood vessels. Notably, an HSPC sub-population expressing α -catulin, which is also expressed by activated migrating endothelial cells⁴⁶, preferentially localized to bone marrow perisinusoidal domains⁴⁷. We show that many cell types which participate in and form the bone marrow HSC niches, share overlapping cell surface markers, including different endothelial, mesenchymal, haematopoietic

and neuronal cells. We reveal for the first time (to our knowledge), distinct sinusoidal sites for leukocyte trafficking, where entry and exit from the circulation occurs and metabolic parameters such as ROS are regulated (summarized in Extended Data Fig. 10). ROS augmentation at these sites is manifested by penetrating plasma, via fenestrated endothelium, probably carrying HSC ROS-inducing factors, found in higher concentration in the blood¹⁷. Our results suggest that prolonged HSC loitering in the peripheral blood might be hazardous for their stemness properties and should be taken into account for clinical mobilization protocols. This is similar to cases of extraphysiologically driven oxygen shock which elevates ROS levels, reducing LTR-HSC capacity in favour of more committed, maturing progenitors⁴⁸. Endothelial cells are exposed to the highest levels of physiological oxygen tension and therefore have developed internal mechanisms to scavenge excessive ROS molecules and rely mainly on glycolysis to avoid ROS production via oxidative phosphorylation²⁸. ROS levels were not altered in BMECs in all the manipulation models (data not shown); however, NAC-mediated rescue of the HSC phenotype was confirmed. Prolonged NAC treatment reduced HSPC egress to the peripheral blood; however, it had no effect on endothelial barrier integrity (Extended Data Fig. 9r–t). These results reveal how NAC pre-treatment increases transplanted HSC bone marrow engraftment^{49,50}, as it has no effect on barrier permeability, allowing successful HSPC bone marrow homing, and it may also promote a ROS-low supportive stem cell retaining microenvironment. The dynamic and versatile endothelial barrier may enable new approaches to control stem cell functions relevant for clinical stem cell mobilization and transplantation protocols to enhance HSPCs egress to the peripheral blood, to promote successful bone marrow lodgment of transplanted HSCs, and to expand engrafting HSCs following transplantation, by restricting or permitting the degree of plasma penetration.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 January 2015; accepted 25 February 2016.

Published online 13 April 2016.

- Rafii, S., Butler, J. M. & Ding, B. S. Angiocrine functions of organ-specific endothelial cells. *Nature* **529**, 316–325 (2016).
- Lapidot, T., Dar, A. & Kollet, O. How do stem cells find their way home? *Blood* **106**, 1901–1910 (2005).
- Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327–334 (2014).
- Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature* **507**, 323–328 (2014).
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Sipkins, D. A. *et al.* In vivo imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435**, 969–973 (2005).
- Colmone, A. *et al.* Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science* **322**, 1861–1865 (2008).
- Lo Celso, C. *et al.* Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457**, 92–96 (2009).
- Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
- Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
- Kunisaki, Y. *et al.* Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* **502**, 637–643 (2013).
- Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4**, 263–274 (2009).
- Isern, J. *et al.* The neural crest is a source of mesenchymal stem cells with specialized hematopoietic stem cell niche function. *Elife* **3**, e03696 (2014).
- Ito, K. *et al.* Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nature Med.* **12**, 446–451 (2006).
- Miyamoto, K. *et al.* Foxo3a is essential for maintenance of the hematopoietic stem cell pool. *Cell Stem Cell* **1**, 101–112 (2007).
- Tesio, M. *et al.* Enhanced c-Met activity promotes G-CSF-induced mobilization of hematopoietic progenitor cells via ROS signaling. *Blood* **117**, 419–428 (2011).

17. Golan, K. *et al.* S1P promotes murine progenitor cell egress and mobilization via S1P1-mediated ROS signaling and SDF-1 release. *Blood* **119**, 2478–2488 (2012).
18. Zhao, M. *et al.* Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nature Med.* **20**, 1321–1326 (2014).
19. Bruns, I. *et al.* Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. *Nature Med.* **20**, 1315–1320 (2014).
20. Nakamura-Ishizu, A., Takubo, K., Fujioka, M. & Suda, T. Megakaryocytes are essential for HSC quiescence through the production of thrombopoietin. *Biochem. Biophys. Res. Commun.* **454**, 353–357 (2014).
21. Nombela-Arrieta, C. *et al.* Quantitative imaging of haematopoietic stem and progenitor cell localization and hypoxic status in the bone marrow microenvironment. *Nature Cell Biol.* **15**, 533–543 (2013).
22. Ono, N. *et al.* Vasculature-associated cells expressing nestin in developing bones encompass early cells in the osteoblast and endothelial lineage. *Dev. Cell* **29**, 330–339 (2014).
23. Ludin, A. *et al.* Monocytes-macrophages that express α -smooth muscle actin preserve primitive hematopoietic cells in the bone marrow. *Nature Immunol.* **13**, 1072–1082 (2012).
24. Yamazaki, S. *et al.* Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. *Cell* **147**, 1146–1158 (2011).
25. Papayannopoulou, T., Priestley, G. V., Nakamoto, B., Zafiropoulos, V. & Scott, L. M. Molecular pathways in bone marrow homing: dominant role of $\alpha_4\beta_1$ over β_2 -integrins and selectins. *Blood* **98**, 2403–2411 (2001).
26. Winkler, I. G. *et al.* Vascular niche E-selectin regulates hematopoietic stem cell dormancy, self renewal and chemoresistance. *Nature Med.* **18**, 1651–1657 (2012).
27. De Bock, K. *et al.* Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* **154**, 651–663 (2013).
28. Vandekeere, S., Dewerchin, M. & Carmeliet, P. Angiogenesis revisited: an overlooked role of endothelial cell metabolism in vessel sprouting. *Microcirculation* **22**, 509–517 (2015).
29. Spencer, J. A. *et al.* Direct measurement of local oxygen concentration in the bone marrow of live animals. *Nature* **508**, 269–273 (2014).
30. Broxmeyer, H. E. *et al.* Rapid mobilization of murine and human hematopoietic stem and progenitor cells with AMD3100, a CXCR4 antagonist. *J. Exp. Med.* **201**, 1307–1318 (2005).
31. Heissig, B. *et al.* Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of Kit-ligand. *Cell* **109**, 625–637 (2002).
32. Dar, A. *et al.* Rapid mobilization of hematopoietic progenitors by AMD3100 and catecholamines is mediated by CXCR4-dependent SDF-1 release from bone marrow stromal cells. *Leukemia* **25**, 1286–1296 (2011).
33. Kobayashi, K. *et al.* Stromal cell-derived factor-1 α /C-X-C chemokine receptor type 4 axis promotes endothelial cell barrier integrity via phosphoinositide 3-kinase and Rac1 activation. *Arterioscler. Thromb. Vasc. Biol.* **34**, 1716–1722 (2014).
34. Itkin, T. *et al.* FGF-2 expands murine hematopoietic stem and progenitor cells via proliferation of stromal cells, c-Kit activation, and CXCL12 down-regulation. *Blood* **120**, 1843–1855 (2012).
35. Zhao, M. *et al.* FGF signaling facilitates postinjury recovery of mouse hematopoietic system. *Blood* **120**, 1831–1842 (2012).
36. Murakami, M. *et al.* The FGF system has a key role in regulating vascular integrity. *J. Clin. Invest.* **118**, 3355–3366 (2008).
37. De Smet, F. *et al.* Fibroblast growth factor signaling affects vascular outgrowth and is required for the maintenance of blood vessel integrity. *Chem. Biol.* **21**, 1310–1317 (2014).
38. Houlihan, D. D. *et al.* Isolation of mouse mesenchymal stem cells on the basis of expression of Sca-1 and PDGFR- α . *Nature Protocols* **7**, 2103–2111 (2012).
39. Wei, J. *et al.* Glucose uptake and Runx2 synergize to orchestrate osteoblast differentiation and bone formation. *Cell* **161**, 1576–1591 (2015).
40. Fujita, K. *et al.* Vitamin E decreases bone mass by stimulating osteoclast fusion. *Nature Med.* **18**, 589–594 (2012).
41. Gur-Cohen, S. *et al.* PAR1 signaling regulates the retention and recruitment of EPCR-expressing bone marrow hematopoietic stem cells. *Nature Med.* **21**, 1307–1317 (2015).
42. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
43. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004).
44. Sugimura, R. *et al.* Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150**, 351–365 (2012).
45. Xie, Y. *et al.* Detection of functional haematopoietic stem cell niche using real-time imaging. *Nature* **457**, 97–101 (2009).
46. Bear, M. D. *et al.* Alpha-Catulin co-localizes with vimentin intermediate filaments and functions in pulmonary vascular endothelial cell migration via ROCK. *J. Cell. Physiol.* <http://dx.doi.org/10.1002/jcp.25185> (2015).
47. Acar, M. *et al.* Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature* **526**, 126–130 (2015).
48. Mantel, C. R. *et al.* Enhancing hematopoietic stem cell transplantation efficacy by mitigating oxygen shock. *Cell* **161**, 1553–1565 (2015).
49. Shen, H. *et al.* An acute negative bystander effect of γ -irradiated recipients on transplanted hematopoietic stem cells. *Blood* **119**, 3629–3637 (2012).
50. Hu, L. *et al.* Antioxidant N-acetyl-L-cysteine increases engraftment of human hematopoietic stem cells in immune-deficient mice. *Blood* **124**, e45–e48 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Karsenty and M. A. Lichtman for fruitful discussions and for critically reviewing the manuscript. We thank S. Méndez-Ferrer and M. Argueta Hernandez for assistance in studies involving MSPCs and nervous system elements. We thank Z. Porat for technical assistance with ImageStream analysis and R. Rotkopf for assistance with statistical data analysis. This study was partially supported by the Ministry of Science, Technology & Space, Israel and the DKFZ, Germany, grants from the Israel Science Foundation (851/13), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine and EU FP7-HEALTH-2010 (CELL-PID #261387) (T.L.). Confocal studies were supported by the European Research Council Advanced Grant 339409, 'AngioBone' (R.H.A.). Intravital multiphoton studies were supported by NIH grants EB017274 and HL100402 (C.P.L. & D.T.S.).

Author Contributions T.I. designed and performed experiments, analysed data and wrote the manuscript; S.G.-C. helped in the design and execution of experiments and analysed data; J.A.S., A.S., and Y.J. designed and performed intravital related experiments, analysed data and helped with writing the manuscript; S.K.R. and A.P.K. designed and performed confocal related experiments and analysed data; G.L., I.M., M.G.P., A.K., A.L., and O.K. helped with experiments; G.S. helped and guided some intravital related experiments; J.M.B. and S.R. helped in design of endothelial-related studies; R.H.A. helped and guided in design of confocal and *in vivo* endothelial related experiments; D.T.S. helped and guided design of *in vivo* intravital live imaging experiments and wrote the manuscript; and T.L. and C.P.L. guided and designed the research and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.L. (tsvee.lapidot@weizmann.ac.il) or C.P.L. (charles_lin@hms.harvard.edu).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment

Animals. C57BL/6 (CD45.2) mice were purchased from Harlan Laboratories (Rehovot, Israel). B6.SJL (CD45.1) mice were bred in-house. Transgenic Ly6a(Sca-1)-EGFP mice and transgenic ROSA26-eYFP (Endo^{YFP}) reporter mice were purchased from Jackson Laboratories. Transgenic nestin-GFP mice were kindly provided by G. N. Enikolopov (Cold Spring Harbour Laboratory, USA). Transgenic c-Kit-EGFP mice were kindly provided by S. Ottolenghi (University of Milano-Bicocca, Italy). Transgenic VE-cadherin (Cdh5, PAC)-CreERT2 mice were kindly provided by R. H. Adams (Max Planck Institute for Molecular Biomedicine, Germany). Conditional mutants carrying *loxP*-flanked *Cxcr4* were provided by D. Scadden (Harvard University, Cambridge, USA). Conditional mutants carrying *loxP*-flanked *Fgfr1* and *Fgfr2* (*Fgfr1/Fgfr2*^{lox/lox}) mice were provided by S. Werner (Institute of Cell Biology, Switzerland) and by D. Ornitz (Washington University School of Medicine, USA). To induce endothelial-specific Cre activity and gene inactivation/expression, adult VE-cadherin (Cdh5, PAC)-CreERT2 mice interbred with *Cxcr4*^{lox/lox} (Endo^{ΔCxcr4}) or *Fgfr1/2*^{lox/lox} (Endo^{ΔFgfr1/2}) or with ROSA26-eYFP mice (Endo^{YFP}) were injected intraperitoneally (i.p.) with Tamoxifen (Sigma, T5648) at 1 mg per mouse per day for 5 days. Mice were allowed to recover for 4 weeks after tamoxifen injections, before euthanasia and experimental analysis. Mice carrying only VE-cadherin (Cdh5, PAC)-CreERT2 transgene or the *Cxcr4*^{lox/lox}/*Fgfr1/2*^{lox/lox} mutations were used as wild-type controls to exclude non-specific effects of Cre activation or of floxed alleles mutation. The endothelial *Fgfr1/2* deletion was confirmed by qRT-PCR measurements of *Cxcr4* and *Fgfr1/2* mRNA from isolated BMECs.

Male and female mice at 8–12 weeks of age were used for all experiments. All mouse offspring from all strains were routinely genotyped using standard PCR protocols. Sample size was limited by ethical considerations and background experience in stem cell transplantation (bone marrow transplantation) which exists in the laboratory for many years and other published manuscripts in the stem cell field, confirming a significant difference between means. No randomization or blinding was used to allocate experimental groups and no animals were excluded from analysis. All mutated or transgenic mouse strains had a C57BL/6 background. All experiments were done with approval from the Weizmann Institute Animal Care and Use Committee. Mice that were maintained at the Weizmann Institute of Science were bred under defined flora conditions. Two-photon *in vivo* microscopy procedures that were performed in Harvard Medical School were approved by the Institutional Animal Care and Use Committee at Massachusetts General Hospital. ***In vivo* treatments.** AMD3100 (Sigma-Aldrich) 5 mg per kg was used to treat mice by subcutaneous (s.c.) injection. Mice were euthanized 30 min later.

Recombinant murine FGF-2 (ProSpec) 200 µg per kg was used to treat mice by i.p. injections for seven consecutive days.

Neutralizing rat anti-VE-cadherin antibodies or rat IgG (eBioscience) at 50 µg per mouse per day were used to treat mice by intravenous (i.v.) injections for 2 or 5 days.

Neutralizing mouse anti-CXCR4 antibodies (12G5 clone) or mouse IgG (eBioscience) at 50 µg per mouse were administered twice, with a 30 min interval, by intravenous (i.v.) injections.

To inhibit ROS production, the antioxidant *N*-acetyl-L-cysteine (NAC; Sigma-Aldrich) was administered by i.p. injection of 130 mg per kg for 2, 5 or 7 days. Mice were euthanized 2–4 h following the final injection.

Immunofluorescence. For standard and confocal fluorescent microscopy, femurs were fixed for 2 h in 4% paraformaldehyde, which was replaced and then the samples were washed with 30% sucrose, embedded in optimum cutting temperature compound, and then snap-frozen in *N*-methylbutane chilled in liquid nitrogen. Sections (5–10 µm) were generated with a CM1850 Cryostat (Leica) at –25 °C with a tungsten carbide blade (Leica) and a CryoJane tape transfer system (Instrumedics), and were mounted on adhesive-coated slides (Leica), fixed in acetone and air-dried. Sections were stained by incubation overnight at 4 °C with primary antibodies, followed by 1 h incubation of secondary antibody at room temperature and in some cases also nuclei labelling by Hoechst 33342 (Molecular Probes) for 5 min at room temperature. Standard analysis (5–6 µm sections) was performed with Olympus BX51 microscope and Olympus DP71 camera. Confocal analysis (10 µm sections) was performed using a Zeiss LSM-710 microscope. In some cases, for BMBV morphological and phenotypical confocal analysis, femurs and tibias were fixed for 2 h in 4% paraformaldehyde, decalcified with 0.5 M EDTA at 4 °C with constant shaking, immersed into 20% sucrose and 2% polyvinylpyrrolidone (PVP) solution for 24 hours, then embedded and frozen in 8% gelatin (porcine) in presence of 20% sucrose and 2% PVP. Sections (80–300 µm) were generated using low-profile blades on a CM3050 cryostat (Leica). Bone

sections were air-dried, permeabilized for 10 min in 0.3% Triton X-100, blocked in 5% donkey serum at room temperature for 30 min, and incubated overnight at 4 °C with primary antibodies. Confocal analysis was performed using a Zeiss LSM-780 microscope. Z-stacks of images were processed and 3D-reconstructed with Imaris software (version 7.00, Bitplane). As previously described⁴, tile scan images were produced by combining the signal of multiple planes along the Z-stalk of bone sections to allow visualization of the distinct types of bone marrow blood vessels and the cells in their surroundings. For the quantifications of blood vessel diameters, a region of 600–700 µm from the growth plate towards the caudal region was selected and diameters for arterial and sinusoidal blood vessels were calculated using ImageJ software on the high-resolution confocal images.

Primary and secondary antibodies and relevant information about them are indicated in Supplementary Table 1.

For *in vivo* ROS detection in bone marrow sections, mice were injected i.p. with hydroethidine (Life Technologies) 10 mg per kg, 30 min before euthanasia. For *in vivo* LDL-uptake detection in bone marrow sections, mice were i.v. injected with Dil-Ac-LDL (BTI) 20 µg per mouse, 4 h before euthanasia. Femurs were immediately collected and processed as mentioned earlier. Bone marrow section analysis for scoring ROS^{high} cells was performed using ImageJ software (Extended Data Fig. 1). Multiple sections (>16 per mouse) were generated and analysed from at least 4 mice per group of experimental procedure, in order to confirm biological repeats of the observed data. In some cases, images were processed to enhance the contrast in order to allow better evaluation of co-localization of cellular borders and markers.

Imaris, Volocity (Perkin Elmer), Photoshop and Illustrator (Adobe) software were used for image processing.

Intravital confocal and multiphoton microscopy. For blood vessel imaging in the calvarium of Sca-1-EGFP and nestin-GFP mice, we used a microscope (Ultima Multiphoton; Prairie Technologies) incorporating a pulsed laser (Mai Tai Ti-sapphire; Newport Corp.). A water-immersed 20× (NA 0.95) or 40× (NA 0.8) objective (Olympus) was used. The excitation wavelength was set at 850–910 nm. For intravital imaging, mice were anaesthetized with 100 mg ketamine, 15 mg xylazine and 2.5 mg acepromazine per kg. During imaging, mice were supplied with oxygen and their core temperature was maintained at 37 °C with a warmed plate. The hair on the skullcap was trimmed and further removed using urea-containing lotion and the scalp was incised at the midline. The skull was then exposed and a small steel plate with a cut-through hole was centred on the frontoparietal suture, glued to the skull using cyanoacrylate-based glue and bolted to the warmed plate. To visualize blood vessels, mice were injected i.v. with 2 µl of a 2 µM non-targeted nanoparticles solution (Qtracker 655, Molecular Probes). In some cases, mice were i.v. injected with Dil-Ac-LDL (BTI) 40 µg per mouse, 2 h before their imaging. We typically scanned a 50 µm-thick volume of tissue at 4 µm Z-steps. Movies and figures based on two-photon microscopy were produced using Volocity software (Perkin Elmer). For live imaging of blood vessels permeability and leukocyte bone marrow trafficking, a previously described experimental procedures and a home built laser-scanning multiphoton imaging system²⁹, were used with some modifications. Anaesthesia was slowly induced in mice via inhalation of a mixture of 1.5–2% isoflurane and O₂. Once induced, the mixture was reduced to 1.35% isoflurane. By making a U-shaped incision on the scalp, calvarial bone was exposed for imaging and 2% methocellulose gel placed on it for refractive index matching.

For bone marrow blood vessel permeability studies, mice were positioned in heated skull stabilization mount which allowed access to the eye for on-stage retro-orbital injection of 40–60 µl of 10 mg ml^{−1} 70 kDa rhodamine-dextran (Life Technologies). Nestin-GFP (excited at 840 nm) and confocal reflectance (at 840 nm) signals were used to determine a region of interest within the mouse calvarial bone marrow for measurement of permeability. Rhodamine-dextran was injected and was continuously recorded (30 frames per second) for the first 10 min after injection. After video acquisition, mice were removed from the microscope and euthanized with CO₂. In some cases, following dextran clearance, the same mice were used for homing experiments to monitor leukocyte cell trafficking in regions and blood vessels that were defined as less or more permeable. For cell homing studies, mice were injected with 2 × 10⁶ DiI-labelled (Life Technologies) lineage depleted immature haematopoietic progenitor cells (Miltenyi depletion) and with 2 × 10⁶ DiI-labelled (Life Technologies) bone marrow MNC isolated from age matched C57BL/6 mice along with 150 µl of 2 nmol per 100 µl Angiosense 750EX (Perkin Elmer) fluorescent blood pool imaging agent, immediately before mounting the mice on a heated stage of a separate confocal/multiphoton microscope. Intravital images of the mouse bone marrow were collected for up to the first 3 h after injection of the cells. After imaging, the mice were removed from the microscope and euthanized with CO₂. Permeability, blood flow/shear rates and homing experiments were repeated, *n* = 3 mice each, measuring multiple blood vessels and events, each mouse regarded as an independent experiment, in order

to confirm biological repeats of the observed data. The contrast and brightness settings of the images in the figures were adjusted for display purposes only.

Permeability and cell homing quantification. For permeability studies, the RGB movies were separated into red (Rhodamine-Dextran), green (nestin-GFP), and blue (reflectance) grayscale image stacks. An image registration algorithm (Normalized Correlation Coefficient, Template Matching) was performed on the red stack using ImageJ (v. 1.47p) to minimize movement artefacts within the image stack. Manual selection of regions of interest (ROI) was performed immediately next to individual vessels within the focus. Permeability of the vessels was calculated using the following equation:

$$P = \frac{V}{A} \times \frac{dI}{dt} \times \frac{1}{I_{in} - I_{out}}$$

P is the permeability of the vessel, V is the volume of the ROI next to the vessel, A is the fractional surface area of the vessel corresponding to the ROI, dI/dt is the intensity of the dye in the ROI as a function of time, I_{in} is the intensity of the dye inside the corresponding vessel at the beginning of measurement, and I_{out} is the intensity of the dye in the ROI at the beginning of measurement. To calculate dI/dt for a given vessel, the change in intensity was measured within the ROI over time and linearly fit the first ~5–40 s of the data. The slope of this linear fit is dI/dt . The ROI intensity curve is only linear for the first 30–40 s, after which it begins to plateau. For cell homing, the number of stationary cells from the calvarial bone marrow images was counted and categorized into two groups: adherent and extravasated. We categorized both cells within the lumen of the vessel and cells in the process of transmigration in the adherent category. Maximum intensity projections of multiple z-stacks of images were used to count the number of cells in the two categories. When there was a gap between cells and vessels in the two-dimensional projection image, those cells were categorized as extravasated. If any part of a cell overlapped a vessel in the projection image, the corresponding three dimensional z-stack was viewed to determine if the cell had undergone extravasation. When it was unclear if a cell had extravasated, it was always categorized as adherent.

Blood flow speed and shear rate quantification. For the flow speed measurement, red blood cells (RBCs) were labelled with 15 μ M CFSE for 12 min at 37°C in PBS supplemented with 1 g per litre of glucose and 0.1% BSA. About 0.6 billion RBCs were injected (i.v.). 40 μ l of rhodamineB-dextran 70 kDa (10 mg ml⁻¹) was retro-orbitally injected immediately before imaging for visualizing bone marrow vasculature. Videos of confocal images of blood vessel (RhodamineB, excitation: 561 nm, emission: 573–613 nm) and labelled RBCs (CFDA-SE, excitation: 491 nm, emission: 509–547 nm) were taken with the speed of 120 frames per second. Individual RBCs were traced over a couple of frames. Total displacement of the RBCs was measured by ImageJ and the speed of blood flow was calculated by:

$$\frac{\text{Total displacement of RBC } (\mu\text{m})}{\text{Time } (= \text{number of frames} \times \frac{1}{120} \text{ sec})}$$

To calculate the shear rate, we assumed that the vessels were straight (straight cylinder) and the blood is an ideal Newtonian fluid with constant viscosity. Under these conditions, the shear rate (du/dr) can be calculated by $du/dr = 8 \times u/d$ (u is the average blood flow speed which was measured by tracing labelled RBCs and d is the diameter of the blood vessel as measured using ImageJ).

Flow cytometry. Immunostaining signal intensity was analysed with MacsQuant (Miltenyi, Germany) or with a FACS LSRII (BD Biosciences) with FACSDiva software, data were analysed with FlowJo (Tree Star). Data of the expression of molecules by cells was analysed and presented as MFI (mean fluorescent intensity). To acquire single bone marrow cell suspensions, freshly isolated bones were cleaned, flushed and crushed using liver digestion medium (LDM, Invitrogen) supplemented with 0.1% DNaseI (Roche) and further digested for 30 min at 37°C, under shaking conditions. Following the incubation time, cells were filtered and washed extensively. To isolate and acquire mononuclear cells (MNC) from the peripheral blood PB, blood was collected from the heart using heparinized syringes and MNC were separated using Lymphoprep (Axis-Shield) according to the manufacturer's instructions. Isolated bone marrow and peripheral blood MNC cells underwent red blood cell lysis (Sigma) before staining. Cells were stained for 30 min at 4°C in standard flow cytometry buffer with primary antibodies and, where indicated, with secondary antibodies. Information about the primary and secondary antibodies can be found in the antibody information (Supplementary Table 1).

For antigens that required intracellular staining, cell surface staining was followed by cell fixation and permeabilization with the Cytofix/Cytoperm kit following the manufacturer's instructions (BD Biosciences). In case of internal GFP labelled cells, cells were fixed for 20 min with 4% PFA at room temperature, washed and incubated at room temperature for 1 h in 30% sucrose. Cells were washed with flow cytometry buffer and further permeabilized. For intracellular

ROS detection, cells were incubated for 10 min at 37°C with 2 μ M hydroethidine (Life Technologies). For glucose uptake detection, cells were incubated for 30 min at 37°C with the glucose analogue 2-NBDG (Life Technologies). For detection of apoptotic cells, cells were resuspended in annexinV binding buffer (BioLegend) and stained with Pacific Blue AnnexinV (BioLegend).

ImageStream analysis. Bone marrow cells were enriched for the lineage negative population, prepared as indicated for flow cytometry and analysed using an ImageStreamX (Amnis) machine. Samples were visualized and analysed for the expression of markers and antigens with IDEAS 4.0 software (Amnis). Single-stained control cells were used to compensate fluorescence between channel images. Cells were gated for single cells with the area and aspect ratio features or, for focused cells, with the Gradient RMS feature. Cells were then gated for the selection of positively stained cells only with their pixel intensity, as set by the cutoff with IgG and secondary antibody control staining. At least 5 samples from 5 mice were analysed to confirm biological repeats of observed data.

Calcitonin and PTH ELISA kit assays. Detection of mouse calcitonin (Cusabio) and mouse PTH (Cloud-Clone Corp.) levels in bone marrow supernatants was performed according to the manufacturer's instructions.

CFU assays. CFU-GM and CFU-F assays were previously described³⁴. For CFU-Ob assay (also known as mineralized nodule formation assay), CFU-F medium was supplemented with 50 μ g ml⁻¹ ascorbic acid and with 10 mM β -glycerophosphate. After 3 weeks, cultures were washed, fixed and stained using Alizarin red for mineralized matrix. The area of mineralized nodules per cultured well was quantified based on image analysis using ImageJ.

In vitro assays. Bone marrow cells were isolated after sterile bone flushing, crushing and digestion (as previously described). After washing, total bone marrow cells were incubated in medium supplemented with or without 25% blood plasma or supplemented with 20 ng ml⁻¹ TGF- β 1 (ProSpec) for 2 h. Plasma was isolated and collected from the upper fraction acquired from the peripheral blood after 5 min centrifugation at 1,500 r.p.m.

In vivo Evans blue dye bone marrow penetration assay. Bone marrow vascular endothelial barrier function was assessed using the Evans Blue Dye (EBD) assay. Evans Blue (Sigma-Aldrich) 20 mg per kg was injected i.v. 4 h before mice were euthanized. In each experiment, a non-injected mouse was used for control blank measurements. Subsequently, mice were perfused with PBS via the left ventricle to remove intravascular dye. Femurs were removed and formamide was used for bone flushing, crushing and chopping. EBD was extracted in formamide by incubation and shaking of flushed and crushed fractions, overnight at 60°C. After 30 min centrifugation at 2,000g, EBD in bone marrow supernatants was quantified by dual-wavelength spectrophotometric analysis at 620 nm and 740 nm. This method corrects the specimen's absorbance at 620 nm for the absorbance of contaminating haem pigments, using the following formula: corrected absorbance at 620 nm = actual absorbance at 620 nm - (1.426(absorbance at 740) + 0.03). Samples were normalized by subtracting control measurements. Levels of EBD bone marrow penetration per femur were expressed as OD₆₂₀/femur and the fold change in EBD bone marrow penetration was calculated by dividing the controls OD₆₂₀/femur from the treated OD₆₂₀/femur in each experiment. Finally, values were normalized per total protein extract as determined by Bradford assay per sample.

Transplantation assays. For competitive LTR assay, B6.SJL (CD45.1) recipient mice were lethally irradiated (1,000 cGy from a caesium source) and injected 5 h later with 2×10^5 donor-derived (C57BL/6 background, CD45.2) bone marrow cells or with 500 μ l of donor-derived whole blood together with 4×10^5 recipient derived (CD45.1) bone marrow cells. Recipient mice were euthanized 24 weeks after transplantation to determine chimaerism levels using flow cytometry analysis. For calculation of competitive repopulating units (CRU), recipient mice were transplanted with limiting dilutions of donor derived bone marrow cells (2.5×10^4 to 2×10^5) together with 2×10^5 recipient derived bone marrow cells. Mice were euthanized after 24 weeks and multi-lineage myelo-lymphoid donor derived contribution in the peripheral blood was assessed using flow cytometry analysis. HSC-CRU frequency and statistical significance was determined using ELDA software (<http://bioinf.wehi.edu.au/software/elda/>).

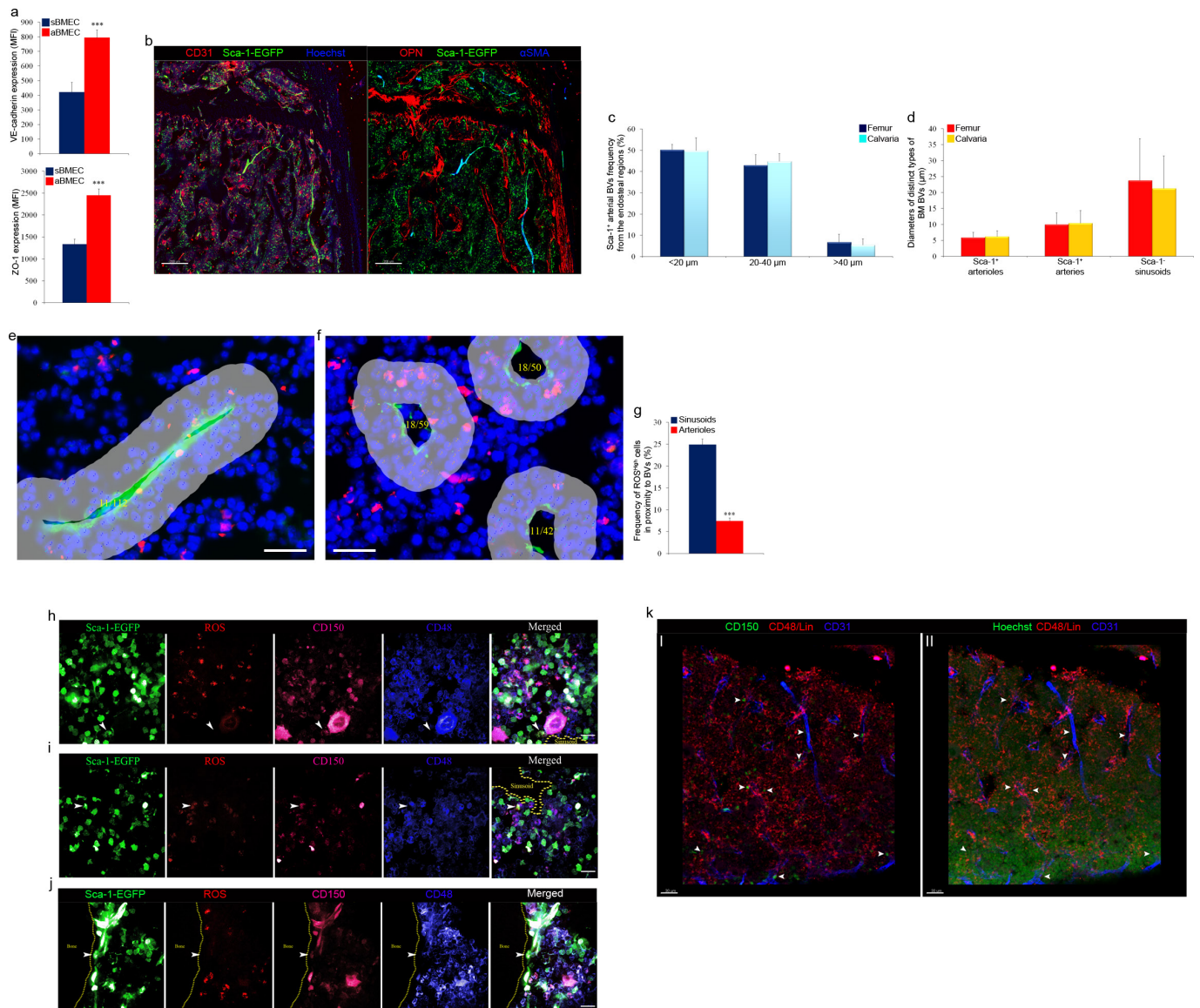
In vivo homing assay. Lineage negative cells were enriched from total bone marrow cells, taken from c-Kit-EGFP mice, using mouse lineage depletion kit (BD) according to the manufacturer's instructions. Non-irradiated recipient mice were transplanted by i.v. injection with 2×10^6 c-Kit-EGFP-labelled Lin⁻ cells. Recipient mice were euthanized 4 h after transplantation. Bone marrow cells were isolated from femurs and stained for flow cytometry as described above. Femur cellularity was determined in order to calculate the number of homed CD34⁺/LSK HSPC per femur.

Quantitative real-time RT-PCR to BMECs. For magnetic isolation of BMECs, freshly recovered bones were processed under sterile conditions as described for BMECs flow cytometry analysis, and post-digestion incubated with biotin rat

anti-mouse CD31 antibodies (BD pharmigen) for 30 min at 4°C. Next, cells were washed and incubated with streptavidin particles plus (BD IMag) for 30 min at 4°C. Positive selection was performed using BD IMagnet (BD) according to the manufacturer's instructions (BD Biosciences). BD IMag buffer (BD) was used for washing and for antibodies dilution. Isolated cells were seeded on fibronectin (Sigma-Aldrich) coated wells and cultured overnight in EBM-2 medium (Lonza) supplemented with EGM-2 SingleQuots (Lonza) at 37°C 5% CO₂. Non-adhesive cells were removed and adherent cells were collected using accutase (eBioscience). Flow cytometry was applied to confirm endothelial identity and >90% purity of recovered cells. BMEC were further processed to isolate RNA. Total RNA was isolated using TRI-Reagent (Sigma-Aldrich) according to the manufacturer's protocol. An aliquot of 2 µg of total RNA was reverse-transcribed using Moloney murine leukaemia virus reverse transcriptase (Promega, Madison, WI) and oligo-dT primers (Promega). Quantitative reverse transcribed-polymerase chain reaction (qRT-PCR) was done using the ABI 7000 machine (Applied Biosystems, Foster City, CA) with SYBR Green PCR Master Mix (Applied Biosystems). Comparative quantization of transcripts was assessed relative to hypoxanthine phosphoribosyl transferase (*Hprt*) levels and amplified with appropriate primers. Primer sequences used were as follows (mouse genes): *Cxcr4*

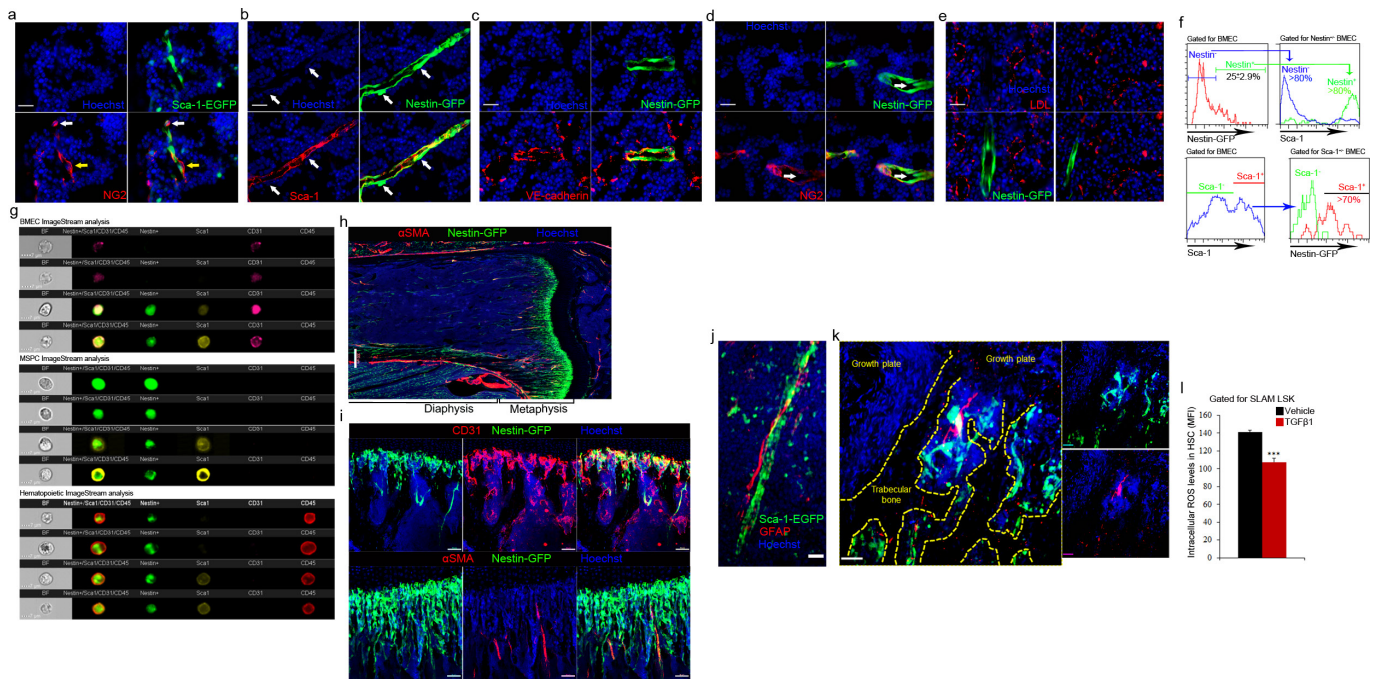
forward 5'-ACGGCTGTAGAGCGAGTGTT-3'; reverse 5'-AGGGTTCC TTGTTGGAGTCA-3'; *Fgfr1* forward 5'-CAACCGTGTGACCAAAGTGG-3'; reverse 5'-TCCGACAGGTCCTTCTCCG-3'; *Fgfr2* forward 5'-ATCCCC CTGCGGAGACA-3'; reverse 5'-GAGGACAGACGCGTTGTTATCC-3'; *Hprt* forward 5'-GCAGTACAGCCCCAAAATGG-3'; reverse 5'-GGTC CTTTTCACCAGCAAGCT-3'.

Statistical analysis. All statistical analyses were conducted with Prism 4.0c version or Excel (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$; NS, not significant). All data are expressed as mean \pm standard error (s.e.m) and all n numbers represent biological repeats. Unless indicated otherwise in figure legends, a Student's two-tailed unpaired t -test was used to determine the significance of the difference between means of two groups. One-way ANOVA or two-way ANOVA was used to compare means among three or more independent groups. Bonferroni post-hoc tests were used to compare all pairs of treatment groups when the overall P value was < 0.05 . A normal distribution of the data was tested using the Kolmogorov-Smirnov test if the sample size allowed. If normal-distribution or equal-variance assumptions were not valid, statistical significance was evaluated using the Mann-Whitney test and the Wilcoxon signed rank test. Animals were randomly assigned to treatment groups. Tested samples were assayed in a blinded fashion.



Extended Data Figure 1 | Properties of distinct types of bone marrow blood vessels and cells in their microenvironment. **a**, Flow cytometry quantitative analysis of VE-cadherin and ZO-1 MFIs by BMEC sub-populations. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; *** $P < 0.005$. **b**, Representative confocal image showing CD31 (red) and Sca-1⁺ (green) arterial blood vessels on proximity to endosteal regions in the metaphysis and representative confocal image of endosteal regions in the metaphysis showing Sca-1⁺ (green) arterial blood vessels, αSMA⁺ (blue) pericytes, and OPN (red) for endosteal borders. Scale bars indicate 200 μm. **c**, Frequencies of Sca-1⁺ arterial blood vessels distribution among zones representing growing distances from the endosteum in the calvaria and femur. **d**, Average diameters of distinct types of blood vessels in the claval and femoral marrow. **e**, **f**, Representative images of arterial blood vessel (green, left) and of sinusoidal blood vessels (green, right) indicating how the frequency of ROS^{high} cells around these blood vessels was scored. The grey-masked areas surrounding the blood vessels indicate the region of distance <20 μm from the blood vessels. Odd numbers (1, 3, and 5) tag the nuclei of cells (blue) found in the region of interest, while even numbers (2, 4, and 6) tag ROS^{high} (red) cells in the region of interest. ImageJ cell counter plugin was used to analyse and score the number of total cells and ROS^{high} cells in the region of interest. Yellow numbers in

the centre of the blood vessels indicate how many ROS^{high} cells are scored out of total cells. Scale bar indicates 20 μm. **g**, Frequency of ROS^{high} cells scored among total bone marrow cells found in proximity (<20 μm) to different bone marrow blood vessels. Mean \pm s.e.m., $n = 24$ bone marrow sections were analysed from $n = 6$ mice. Two-tailed Student's t -test; *** $P < 0.005$. **h–k**, White arrowheads indicate for SLAM HSPC **h**, Representative confocal images with ROS probe (red) of ROS^{low/-}, CD150⁺ (pink)/CD48⁻ (blue) SLAM HSPC, found away (>20 μm) from Sca-1⁺ (green) endosteal blood vessels, neighbouring a megakaryocyte. Yellow dashed line indicates sinusoidal borders. Scale bar indicates 20 μm. **i**, Representative confocal images of ROS^{high} (red), CD150⁺ (pink) and CD48⁻ (blue) SLAM-HSPC, found away (>20 μm) from Sca-1⁺ (green) endosteal blood vessels, surrounded by mature haematopoietic cells. Yellow dashed line indicates sinusoidal borders. Scale bar indicates 20 μm. **j**, Representative confocal images of cells with ROS^{high} (red) levels among CD150⁺ (pink) and CD48⁻ (blue) SLAM-HSPC neighbouring (<20 μm) Sca-1⁺ (green) the endosteal arteriole. Scale bar indicates 20 μm. **k**, Representative tile scan confocal images of bone marrow merged Z-stack showing (I) CD31⁺ blood vessels (blue) and their neighbouring CD150⁺ (green) CD48/Lin (red) negative SLAM HSPC. (II) Cells nuclei are visualized (green) together with CD48/Lin (red) and CD31⁺ blood vessels (blue). Scale bars indicate 30 μm.

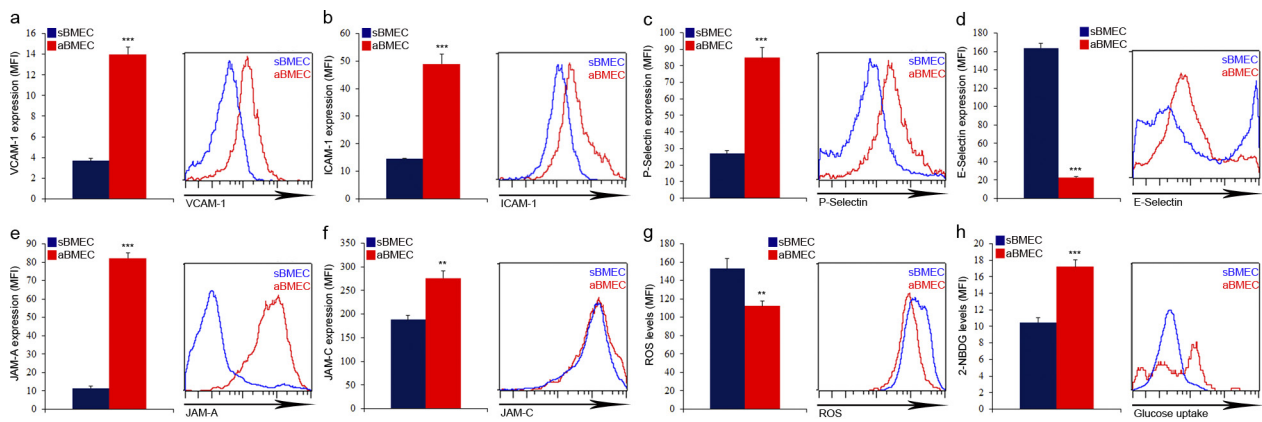


Extended Data Figure 2 | Different populations of nestin-expressing bone marrow cells are associated with nestin-expressing arterioles.

a, Representative fluorescence images of Sca-1⁺ (green) blood vessels and their neighbouring NG2⁺ (red) MSPCs. NG2⁺ MSPCs were either negative (yellow arrow) or positive (white arrow) for Sca-1 expression. Scale bar indicates 20 μ m. **b**, Representative fluorescence images of Sca-1⁺ (red) blood vessels and nestin-GFP labelling (green) blood vessels and MSPCs (white arrows). Scale bar indicates 20 μ m. **c**, Representative fluorescence images of nestin⁺ (green) blood vessels and VE-cadherin (red) staining, showing that nestin⁺ blood vessel structures are co-stained with VE-cadherin while neighbouring sinusoids are VE-cadherin⁺/nestin⁻. Scale bar indicates 20 μ m. **d**, Representative fluorescence images of nestin⁺ (green) blood vessels and their neighbouring NG2⁺ (red) MSPCs. NG2⁺/nestin⁺ MSPCs surrounded NG2⁻/nestin⁺ aBMECs with elongated nuclei (white arrow). Scale bar indicates 20 μ m. **e**, Representative fluorescence images of large- and small-diameter nestin⁺ (green) blood vessels and blood vessels positive for LDL (red) uptake, indicating that nestin⁺ labels arteries and arterioles but not sinusoids. Scale bar indicates 20 μ m. **f**, Representative flow cytometry histogram plots for gated BMECs, showing nestin-GFP expression on BMEC subpopulation which is Sca-1⁺ or nestin-GFP expression by Sca-1^{+/+} BMEC subpopulation.

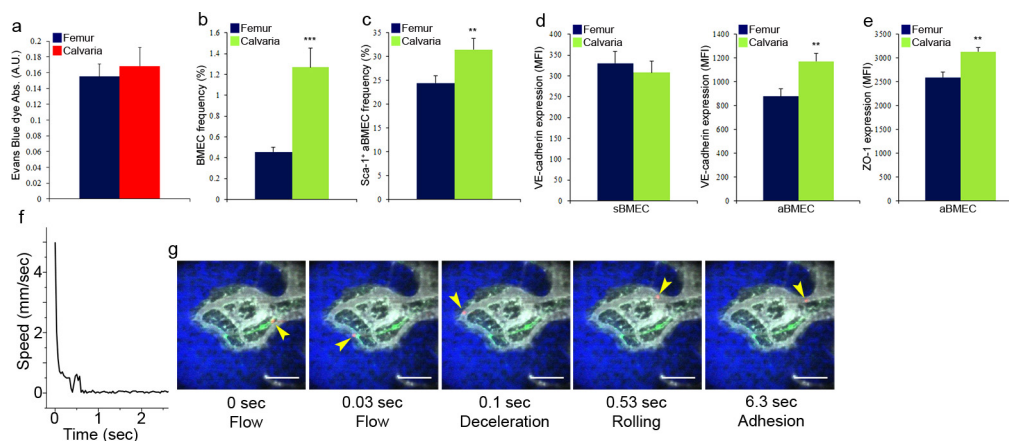
Mean \pm s.e.m., $n = 6$ mice from three independent experiments.

g, Representative ImageStream images of CD45⁻CD31⁺Sca-1⁺nestin⁻sBMECs and CD45⁻CD31⁺Sca-1⁺nestin⁺aBMECs, CD45⁻CD31⁻Sca-1^{+/+}nestin⁺MSPCs, and CD45⁺CD31⁻Sca-1^{+/+}nestin⁺haematopoietic cells. **h**, Representative confocal tile scan of nestin-GFP (green) femur stained with α SMA (red). Scale bar indicates 200 μ m. **i**, Representative confocal images of endosteal regions in the metaphysis showing α SMA (red) enwrapped nestin⁺ (green) and CD31⁺ (white) arterial blood vessels branching into smaller endosteal nestin⁺CD31⁺ arterioles which are not associated with α SMA⁺ pericytes. Endosteal nestin⁺ blood vessels are surrounded by nestin⁺ MSPCs. Scale bars indicate 50 μ m. **j, k**, Representative confocal images of diaphyseal area (j) and metaphyseal area (k) showing GFAP (red, Schwann cell marker) fibres associated with Sca-1⁺ (green) arterial blood vessel (j) or with Sca-1⁺ endosteal arterioles (k). Scale bar indicates 50 μ m (j) and 100 μ m (k). **l**, Bone marrow cells were incubated with 20 ng ml⁻¹ TGF β 1 or vehicle for 2 h. ROS MFI levels in bone marrow SLAM HSPCs were determined by flow cytometry quantitative analysis. Mean \pm s.e.m., $n = 9$ repeats in triplicates from three independent experiments. Two-tailed Student's t -test; *** $P < 0.005$.



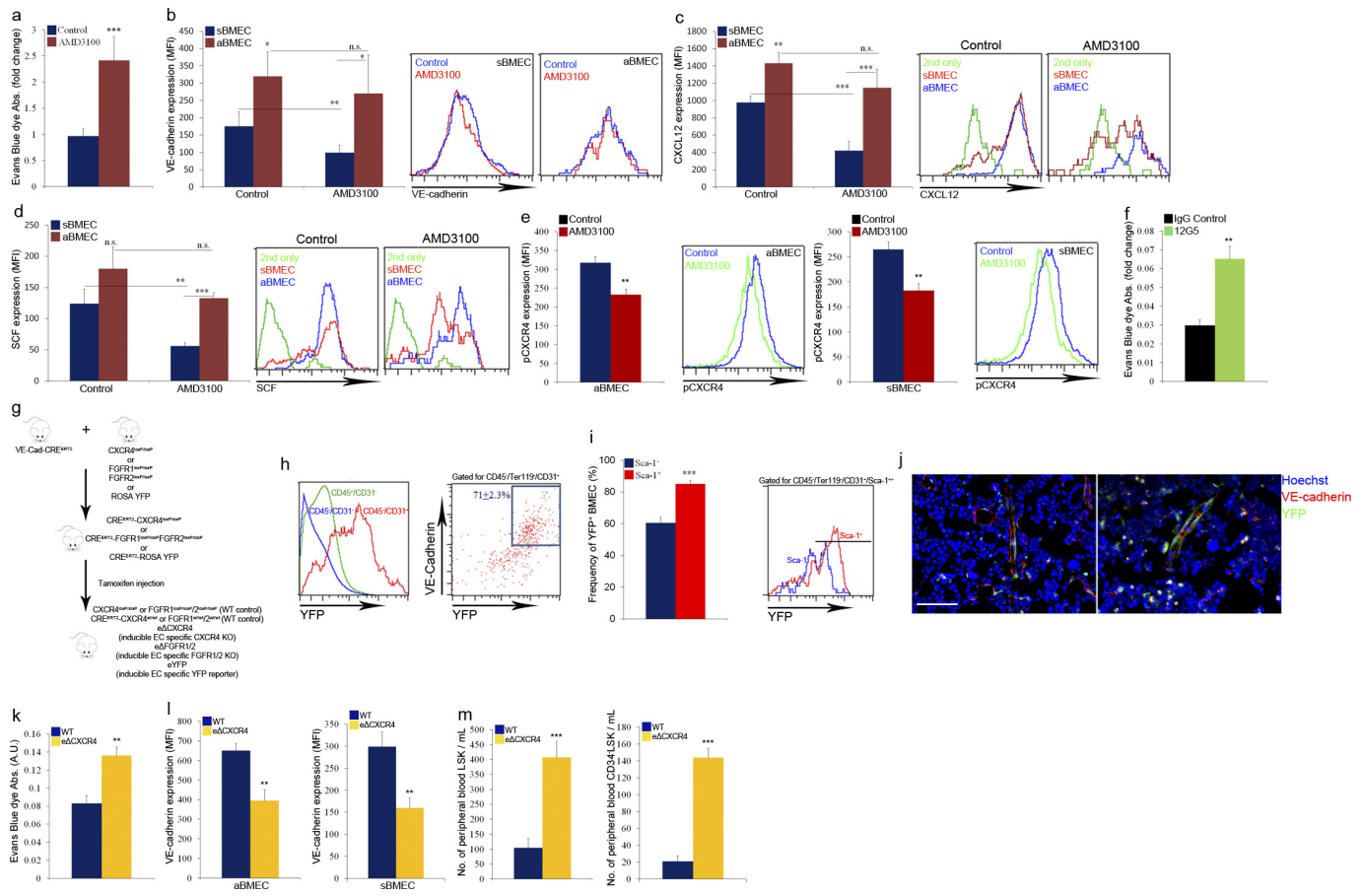
Extended Data Figure 3 | Expression pattern of molecules involved in cellular trafficking by distinct types of blood vessels. a–h, Expression levels (MFI) of indicated surface or intracellular molecules by distinct

types of BMEC as measured by flow cytometry analysis. Mean \pm s.e.m., $n = 8$ Sca-1-EGFP and wild-type mice from two independent experiments). Two-tailed Student's t -test; ** $P < 0.01$, *** $P < 0.005$.



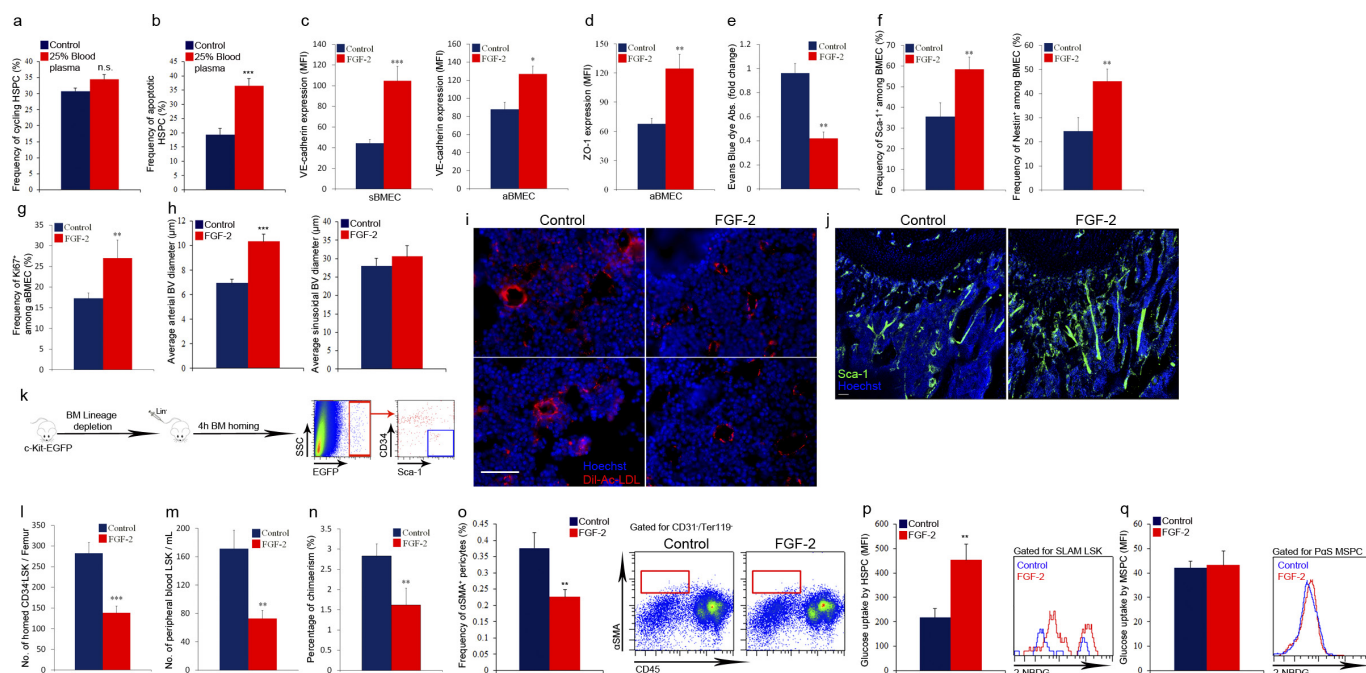
Extended Data Figure 4 | Femoral and calvarial comparison and monitoring of calvarial trafficking. **a**, Evans blue dye (EBD) absorbance following extraction from the femurs or calvarias, was measured using spectrophotometric analysis at 620 nm and 740 nm and normalized to total protein content per femur (Bradford). Mean \pm s.e.m., $n = 6$ mice from two independent experiments. **b–e**, Mean \pm s.e.m., $n = 8$ mice from two independent experiments. Two-tailed Student's t -test; $**P < 0.01$, $***P < 0.005$. **b**, Total BMEC frequency as determined by flow cytometry analysis. **c**, Sca-1⁺ aBMEC frequency as determined by flow cytometry analysis. **d**, **e**, VE-cadherin and ZO-1 expression (MFI) on distinct types

of BMECs as determined by flow cytometry analysis. **f**, A representative plot showing the flow speed of an HSPC passing through a network of nestin-GFP^{+/−} blood vessels as a function of time. Note that the cell temporarily stops within a sinus at ~ 0.4 s and slowly rolls until it adheres again at ~ 0.7 s. **g**, Snapshot images from 0, 0.03, 0.10, 0.53, and 6.3 s taken from Supplementary Video 5. Nestin-GFP (green), HSPC (red), blood vessels (grey), and bone (blue) are displayed. The cell is overlaid on the pre-acquired nestin-GFP, blood vessels, and bone images. Yellow arrows indicate for the location of the trafficking HSPC. Scale bars indicate 100 μ m.



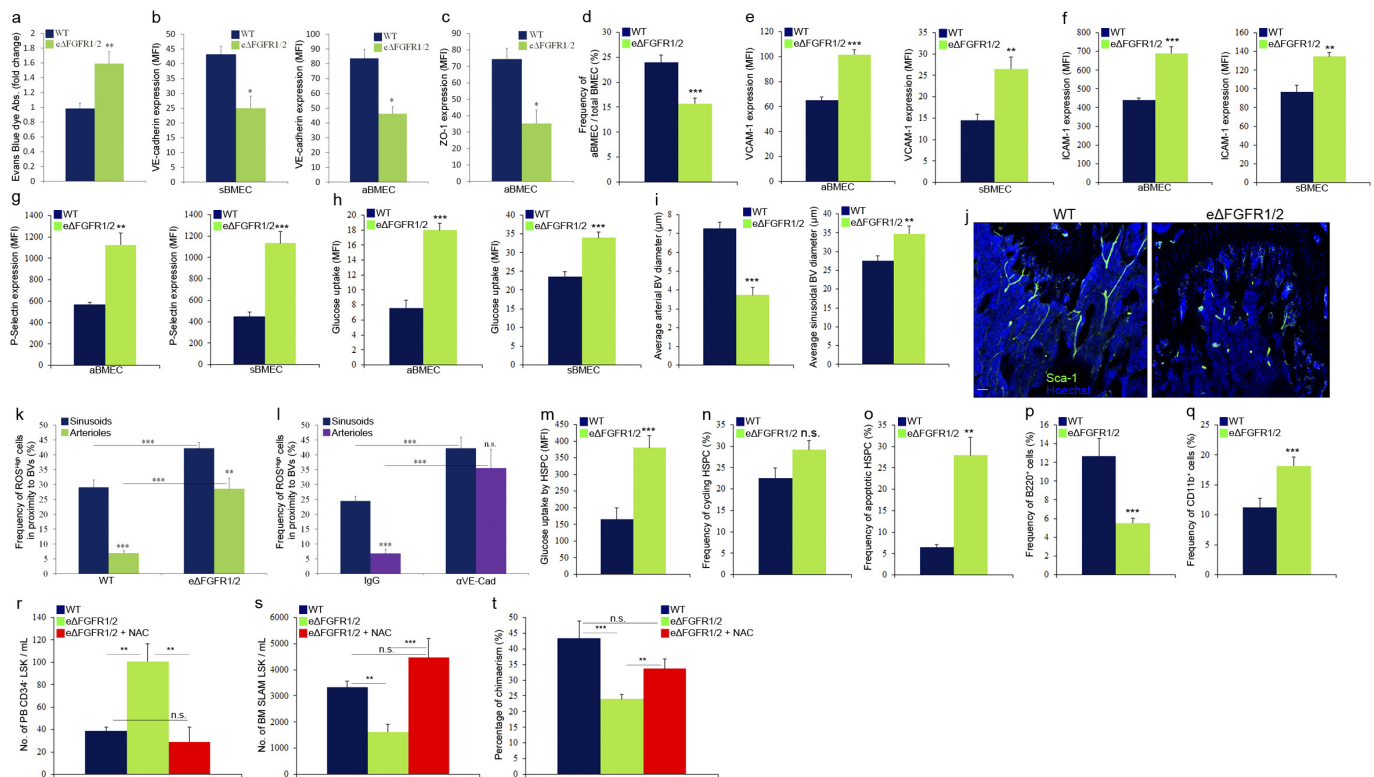
Extended Data Figure 5 | Properties of distinct types of bone marrow blood vessels under HSPC mobilization conditions and the role of the endothelial CXCL12–CXCR4 axis. **a–e**, C57BL/6 or nestin–GFP mice received a single injection of AMD3100 (5 mg per kg) and were analysed 5 min (for pCXCR4) or 30 min later. Mean \pm s.e.m., $n = 7$ mice from three independent experiments. Two-way ANOVA with Bonferroni's multiple comparison post-hoc test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$. **a**, Evans blue dye (EBD) absorbance following EBD (30 mg per kg) injection together with AMD3100. **b–d**, Flow cytometry quantitative analysis and representative histogram plots of VE-cadherin, membrane-bound CXCL12, and membranal SCF MFIs. **e**, Intracellular CXCR4 phosphorylation (pCXCR4) levels (MFI) in distinct types of BMECs as measured by flow cytometry analysis and representative histogram plots. **f**, C57BL/6 mice received two injections (30 min interval) of 50 μ g 12G5 CXCR4 neutralizing antibodies or IgG control followed by EBD injection. EBD absorbance following extraction from the femur was measured using spectrophotometric analysis at 620 nm and 740 nm. Mean \pm s.e.m., $n = 6$ mice from two independent experiments. Two-tailed Student's t -test, ** $P < 0.01$. **g**, Endothelial cell (EC)-specific inducible deletion of *Cxcr4* ($Endo^{\Delta Cxcr4}$) or *Fgfr1/2* ($Endo^{\Delta Fgfr1/2}$) in mice. Mice harbouring *loxP* sites flanking *Cxcr4* or *Fgfr1* and *Fgfr2* genes were crossed with a mouse line with endothelial-cell-specific VE-cadherin promoter-driven CreERT2 (VE-cadherin (Cdh5, PAC)-CreERT2). Specificity of VE-cadherin (Cdh5, PAC)-CreERT2 was validated in reporter mice carrying enhanced

YFP protein following floxed stop codon ($Endo^{YFP}$). *Cxcr4* or *Fgfr1/2* deletion or YFP expression in endothelial cells was induced by tamoxifen injection. Mouse analysis was performed 4 weeks after tamoxifen-induced Cre activity. Mice carrying only the *Cxcr4*^{lox/lox} or *Fgfr1/2*^{lox/lox} mutations or VE-cadherin (Cdh5, PAC)-CreERT2 transgene served as controls. **h–j**, Mean \pm s.e.m., $n = 12$ mice from four independent experiments. Two-tailed Student's t -test; *** $P < 0.005$. **h**, Representative flow cytometry histogram and dot plots confirming BMEC specific induction of Cre activity by exclusive expression of YFP in ~70% of BMEC. **i**, Frequency of YFP expression and representative histogram plot, among BMEC sub-populations, was determined by flow cytometry quantitative analysis 4 weeks after tamoxifen induction of Cre activity. Note higher Cre activity, indicated by higher YFP signal, in aBMECs. Black line indicates for a positive signal region. **j**, Fluorescent representative images of YFP expression by distinct BMBVs (sinusoids and arteries). **k–m**, Tamoxifen-treated wild-type and $Endo^{\Delta Cxcr4}$ mice were allowed to recover for 4 weeks before studies. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$. **k**, EBD absorbance in wild-type and $Endo^{\Delta Cxcr4}$ mice. **l**, Flow cytometry quantitative analysis of VE-cadherin MFI on BMEC from wild-type and $Endo^{\Delta Cxcr4}$ mice. **m**, Flow cytometry quantitative analysis of blood LSK HSPCs and CD34⁺ LSK HSPCs of wild-type or $Endo^{\Delta Cxcr4}$ mice.



Extended Data Figure 6 | FGF-2 administration remodels the bone marrow vasculature and the stromal compartment while retaining HSPCs in the bone marrow. **a, b**, Bone marrow cells were incubated for 2 h with (25% blood plasma) or without (control) peripheral blood plasma. Mean \pm s.e.m., $n = 9$ repeats from three independent experiments. Two-tailed Student's t -test; *** $P < 0.005$. **a**, Frequencies of cycling Ki67⁺ SLAM LSK HSPC. **b**, Frequencies of apoptotic annexinV⁺ SLAM LSK HSPC. **a–f**, C57BL/6 or nestin-GFP mice were treated with FGF-2 (200 μ g per kg) for 7 days. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; * $P < 0.05$, *** $P < 0.005$, ** $P < 0.01$. **c, d**, Quantitative analysis of VE-cadherin and ZO-1 MFIs on BMECs. **e**, EBD absorbance. **f, g**, Flow cytometry quantitative analysis of BMEC frequencies expressing Sca-1, nestin and intracellular Ki67 cell cycling markers. **h**, Diameters of distinct types of bone marrow blood vessels in the metaphysis region as determined by ImageJ software analysis of high-resolution confocal images. **i**, Fluorescent representative images of LDL (red) uptake by sinusoidal BMEC and other bone marrow cells following diffusion into the parenchymal marrow. Note lower LDL uptake and diffusion following FGF-2 treatment. Scale bar indicates 20 μ m.

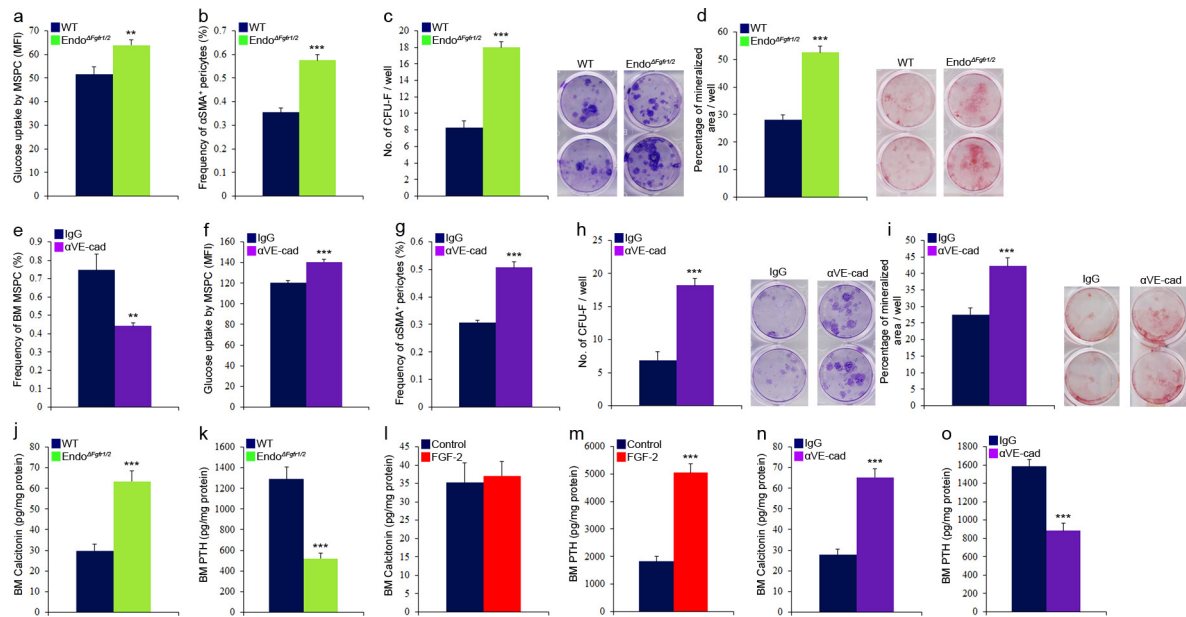
j, Representative confocal images of Sca-1⁺ (green) arterial blood vessels in the metaphysis region. Note higher abundance of arterial blood vessels following FGF-2 treatment. Scale bar indicates 200 μ m. **k**, For the homing assay, bone marrow cells from c-Kit-EGFP labelled mice were lineage depleted, and transplanted to the indicated recipient mice. Four hours after transplantation, bones from recipient mice were recovered, flushed and crushed, and the numbers of homed Lin[−] c-Kit-EGFP⁺ Sca-1⁺ CD34[−] HSPCs were determined per femur by flow cytometry quantitative analysis. **l–q**, Mice were treated with FGF-2 (200 μ g per kg) for 7 days. Mean \pm s.e.m., unless indicated otherwise $n = 12$ mice from three independent experiments. Two-tailed Student's t -test; ** $P < 0.01$, *** $P < 0.005$. **l**, HSPC homing per femur. Mean \pm s.e.m., $n = 8$ mice from three independent experiments. **m**, Numbers of LSK HSPCs in the blood. **n**, Levels of chimaerism indicating LTR-HSC contribution from blood transplant. Mean \pm s.e.m., $n = 20$ mice from two independent experiments. **o**, Frequencies and representative density plots of bone marrow α SMA⁺ pericytes as determined by flow cytometry analysis. **p, q**, Expression levels (MFI) and representative histograms of glucose uptake by HSPCs and MSCs, respectively, were determined by flow cytometry analysis.



Extended Data Figure 7 | Genetic breaching of the endothelial barrier remodels the bone marrow vasculature and the stromal compartment while enhancing HSPC egress in a ROS dependent manner.

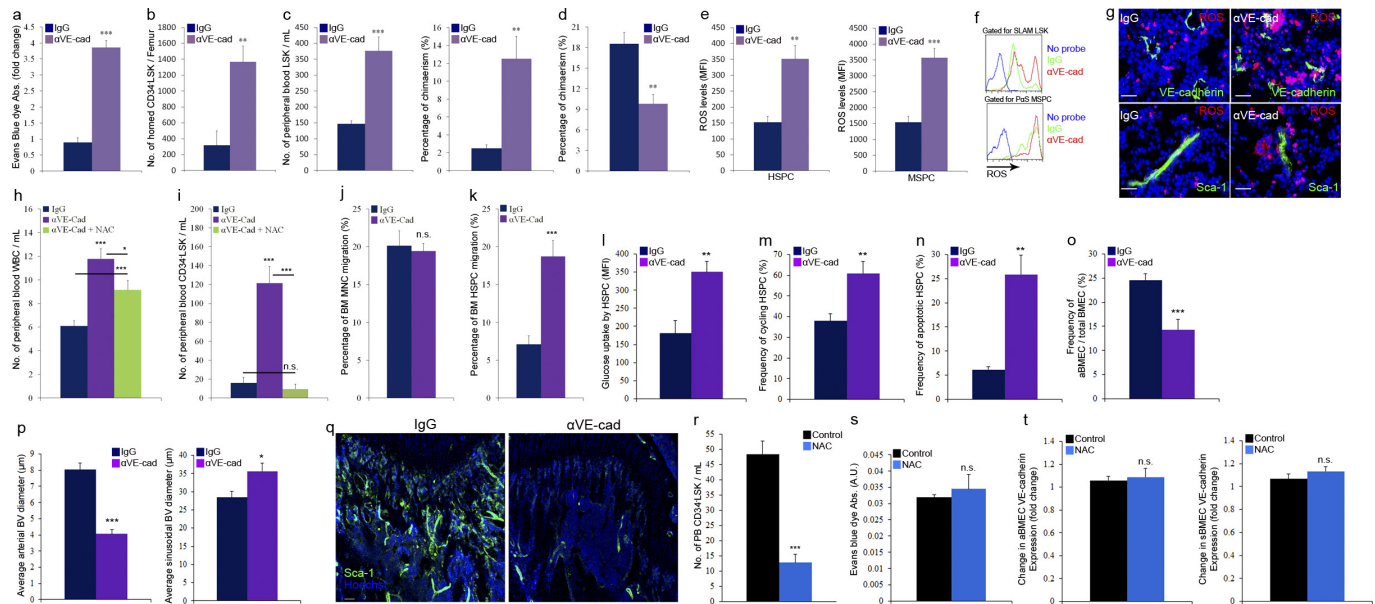
a, EBD absorbance. Mean \pm s.e.m., $n = 6$ mice from three independent experiments. Two-tailed Student's t -test; $**P < 0.01$. **b, c**, Quantitative analysis of VE-cadherin and ZO-1 MFIs on BMEC. Mean \pm s.e.m., $n = 6$ mice from three independent experiments. Two-tailed Student's t -test; $*P < 0.05$. **d–h**, Flow cytometry quantitative analysis of BMEC frequencies, surface and intracellular molecules expression (MFI), in wild-type or Endo $\Delta Fgfr1/2$ mice. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; $**P < 0.01$, $***P < 0.005$. **i**, Diameters of distinct types of bone marrow blood vessels in the metaphysis region as determined by ImageJ software analysis of high-resolution confocal images. **j**, Representative confocal representative images of Sca-1⁺ (green) arterial blood vessels in the metaphysis region. Note lower abundance of arterial blood vessels in Endo $\Delta Fgfr1/2$ mice. Scale bar indicates 200 μ m. **k, l**, Mean \pm s.e.m., $n = 16$ bone marrow sections were analysed from $n = 4$ mice. Two-way ANOVA with Bonferroni's multiple comparison post-hoc test; $**P < 0.01$, $***P < 0.005$. **k**, Frequency of ROS^{high} cells scored among total bone marrow cells found in proximity ($< 20 \mu$ m) to different bone marrow blood vessels, in wild-type or Endo $\Delta Fgfr1/2$ mice. **l**, Frequency of ROS^{high} cells scored

among total bone marrow cells found in proximity ($< 20 \mu$ m) to different bone marrow blood vessels, in C57BL/6 mice treated with neutralizing rat anti-VE-cadherin antibodies or rat IgG control antibodies (50 μ g per mouse per day) for 2 days. **m–o**, Flow cytometry quantitative analysis of HSPC glucose uptake (**m**) (MFI), frequency of cycling HSPC (**n**) and apoptotic HSPC (**o**). Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; $**P < 0.01$, $***P < 0.005$. **p, q**, Frequencies of donor-derived lymphoid B220⁺ or myeloid CD11b⁺ cells in the PB of recipient mice, as were determined 24 weeks after transplantation by flow cytometry. Mean \pm s.e.m., $n = 18$ donor mice from two independent experiments, for 3 recipient mice per donor. Two-tailed Student's t -test; $***P < 0.005$. **r–t**, Wild-type or Endo $\Delta Fgfr1/2$ mice were treated with NAC (130 mg per kg) or PBS for 7 days. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; $**P < 0.01$, $***P < 0.005$. **r**, Number of circulating peripheral blood HSPC as determined by quantitative flow cytometry analysis. **s**, Number of bone marrow SLAMF LSK HSPC as determined by quantitative flow cytometry analysis. **t**, Levels of chimerism, indicating LTR-HSC contribution, were determined 24 weeks after transplantation by flow cytometry ratio analysis (CD45.2/(CD45.2 + CD45.1)). Mean \pm s.e.m., $n = 24$ donor mice from two independent experiments, for 3 recipient mice per donor.



Extended Data Figure 8 | Endothelial barrier manipulation affects stromal properties, development and the levels of bone remodelling hormones. **a–d**, Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; $**P < 0.01$, $***P < 0.005$. **a**, Glucose uptake by bone marrow MSC as determined by quantitative flow cytometry (MFI) analysis. **b**, Frequencies of bone marrow α SMA $^{+}$ pericytes as determined by flow cytometry analysis. **c**, Average number of scored (ImageJ) CFU-F per well and representative images. **d**, Average determined (ImageJ) percentage of mineralized area per well and representative images. **e–i**, C57BL/6 mice were treated with neutralizing rat anti-VE-cadherin antibodies or rat IgG control antibodies (50 μ g per mouse per day) for 5 days. Mean \pm s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test and one-way ANOVA with Bonferroni's multiple comparison post-hoc test; $**P < 0.01$,

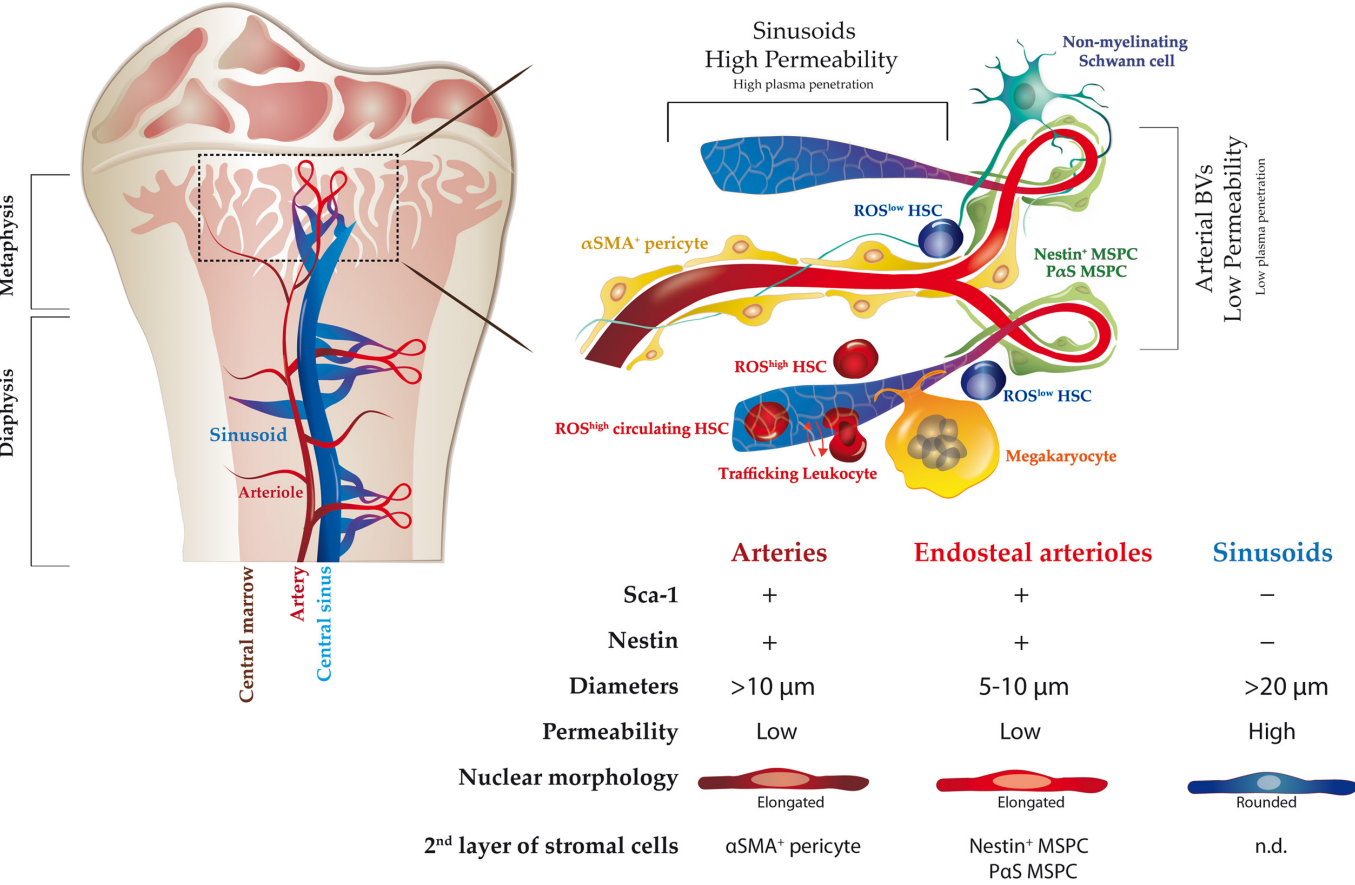
$***P < 0.005$. **e**, Frequency of bone marrow MSC as determined by flow cytometry quantitative analysis. **f**, Glucose uptake by bone marrow MSC as determined by quantitative flow cytometry (MFI) analysis. **g**, Frequencies of bone marrow α SMA $^{+}$ pericytes as determined by flow cytometry analysis. **h**, Average number of scored (ImageJ) CFU-F per well and representative images. **i**, Average determined (ImageJ) percentage of mineralized area per well and representative images. **j–o**, Bone marrow supernatants from wild-type or $Endo^{\Delta Fgf1/2}$, PBS or FGF-2 (200 μ g per kg) treated for 7 days, and IgG or rat anti-VE-cadherin (50 μ g per mouse per day) for 5 days, were isolated and bone marrow concentrations of calcitonin and PTH hormones were determined using an ELISA assay. Mean \pm s.e.m., $n = 9$ mice per group from three independent experiments. Two-tailed Student's t -test; $***P < 0.005$.



Extended Data Figure 9 | Pharmacological breaching of the endothelial barrier remodels the bone marrow vasculature and the stromal compartment while enhancing HSPCs egress in a ROS-dependent manner. **a–j**, C57BL/6 mice were treated with neutralizing rat anti-VE-cadherin antibodies or rat IgG control antibodies (50 μg per mouse per day) for 5 days. Mean ± s.e.m., unless otherwise indicated, $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; $*P < 0.05$, $**P < 0.01$, $***P < 0.005$. **a**, EBD absorbance. **b**, HSPCs homing per femur. **c**, Quantitative analysis of blood LSK HSPC and chimaerism levels indicating LTR-HSC contribution. Mean ± s.e.m., $n = 10$ mice from two independent experiments. **d**, Chimaerism levels indicating LTR-HSC contribution. Mean ± s.e.m., $n = 18$ donor mice from two independent experiments, for 3 recipient mice per donor. **e**, **f**, Quantitative analysis and representative histogram plots of HSPCs and P α S MSPCs ROS MFI. **g**, Representative images of ROS^{high} (red) cells in proximity to blood vessels. Scale bar indicates 20 μm. **h–k**, C57BL/6 mice were treated with neutralizing rat anti-VE-cadherin antibodies or rat IgG control antibodies (50 μg per mouse per day) for 2 days. Where indicated, mice were also treated with NAC (130 mg per kg) or PBS for 2 days. Mean ± s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test and one-way ANOVA with Bonferroni's multiple comparison post-hoc test; $*P < 0.05$, $**P < 0.01$, $***P < 0.005$. **h**, White blood cell (WBC) numbers in the blood circulation were determined using a haematocytometer and Turk lysis of erythrocytes. **i**, Flow cytometry quantitative analysis of CD34⁺ LSK HSPC in the blood circulation. **j–k**, Bone marrow MNC or bone marrow lineage depleted cells from treated mice were seeded on a 5 μm pore transwell

and allowed to migrate for 2 h towards CXCL12 (125 ng ml⁻¹). Following migration, the frequency of migrated bone marrow MNC or CD34⁺/LSK HSPC was determined by flow cytometry quantitative analysis. Note preferential HSPC enhanced migration after VE-cadherin neutralization. **l–p**, C57BL/6 mice were treated with neutralizing rat anti-VE-cadherin antibodies or rat IgG control antibodies (50 μg per mouse per day) for 5 days. Mean ± s.e.m., $n = 9$ mice from three independent experiments. Two-tailed Student's t -test; $**P < 0.01$, $***P < 0.005$. **l**, Glucose uptake levels (MFI) by HSPC as determined by flow cytometry quantitative analysis. **m–n**, Frequencies of Ki67⁺ cycling and AnnexinV⁺ SLAM LSK HSPC as determined by flow cytometry quantitative analysis. **o**, Frequency of Sca-1⁺ aBMEC out of the total as determined by flow cytometry quantitative analysis. **p**, Diameters of distinct types of bone marrow blood vessels in the metaphysis region as determined by ImageJ software analysis of high-resolution confocal images. **q**, Confocal representative images of Sca-1⁺ (green) arterial blood vessels in the metaphysis region. Note lower abundance of arterial blood vessels following 5 days of anti-VE-cadherin treatment. Scale bar indicates 200 μm. **r–t**, C57BL/6 mice were treated with NAC (130 mg per kg) or PBS for 7 days. Mean ± s.e.m., $n = 8$ mice from two independent experiments. Two-tailed Student's t -test; $***P < 0.005$. **r**, Flow cytometry quantitative analysis of CD34⁺ LSK HSPC in the blood circulation. **s**, EBD absorbance following extraction from the femur was measured using spectrophotometric analysis at 620 nm and 740 nm. **t**, VE-cadherin expression levels (MFI) on distinct types of BMEC as determined by flow cytometry quantitative analysis (arterial and sinusoidal respectively).

Distinct bone marrow blood vessels differentially regulate hematopoiesis



Extended Data Figure 10 | Illustration of proposed bone marrow blood vessels model and regulation of haematopoiesis. Bone marrow vasculature is composed of two main types of blood vessels which are arterial blood vessels and sinusoids. Blood enters the bone marrow via the arteries, branching to smaller arterioles, which in proximity to endosteal areas, further branch into small-diameter endosteal arterioles. These endosteal arterioles reconnect to downstream sinusoids which drain the blood into the central sinus and out of the bone marrow. Arterial BMEC have elongated nuclear morphology, express Sca-1 and nestin markers, and display high barrier integrity properties. In addition, arterial blood vessels display the highest blood flow speed and shear rate. Arterial BMEC maintain a microenvironment that promotes low ROS state of HSCs in its surrounding. The second layer of cells associated with arteries is composed of αSMA⁺ pericytes, while endosteal arterioles are associated with HSC-supportive MSPCs. The association of MSPCs and ROS^{low} HSCs with endosteal capillaries suggests the existence of an osteo-vascular niche where the residing HSCs are influenced by both endosteal and vascular elements simultaneously. Innervating Schwann cell nerve fibres, shown

to maintain HSC dormancy, were found to be associated with arteries and endosteal arterioles. More permeable fenestrated sinusoids induce higher ROS state in their surroundings, and have slower internal blood flow, all of which makes them the ultimate candidate to serve as the site for bone marrow cellular trafficking. Megakaryocytes found in sinusoidal sites support and maintain HSPC in a ROS low state. Live real-time imaging indicates that all leukocyte trafficking occurs exclusively via sinusoids. Furthermore, experimental systems manipulating endothelial barrier integrity provide evidence that more fenestrated endothelial state promotes trafficking at the expense of stem cell maintenance. Yet, conditions enhancing endothelial integrity, reducing cellular trafficking promote bone marrow stem cell expansion and maintenance. Peripheral blood plasma, which can penetrate into the bone marrow more easily via fenestrated blood vessels, enhances HSPC migratory capacity but hampers their long-term repopulation capacity and survival. Thus, the state of the endothelial blood–bone-marrow barrier in distinct blood vessels and under steady state or ‘stress’ conditions may have a strong regulatory impact on tissue residing stem cells.

DNA methylation on N^6 -adenine in mammalian embryonic stem cells

Tao P. Wu¹, Tao Wang¹, Matthew G. Seetin², Yongquan Lai³, Shijia Zhu⁴, Kaixuan Lin¹, Yifei Liu¹, Stephanie D. Byrum⁵, Samuel G. Mackintosh⁵, Mei Zhong⁶, Alan Tackett⁵, Guilin Wang⁷, Lawrence S. Hon², Gang Fang⁴, James A. Swenberg³ & Andrew Z. Xiao¹

It has been widely accepted that 5-methylcytosine is the only form of DNA methylation in mammalian genomes. Here we identify N^6 -methyladenine as another form of DNA modification in mouse embryonic stem cells. *Alkbh1* encodes a demethylase for N^6 -methyladenine. An increase of N^6 -methyladenine levels in *Alkbh1*-deficient cells leads to transcriptional silencing. N^6 -methyladenine deposition is inversely correlated with the evolutionary age of LINE-1 transposons; its deposition is strongly enriched at young (<1.5 million years old) but not old (>6 million years old) L1 elements. The deposition of N^6 -methyladenine correlates with epigenetic silencing of such LINE-1 transposons, together with their neighbouring enhancers and genes, thereby resisting the gene activation signals during embryonic stem cell differentiation. As young full-length LINE-1 transposons are strongly enriched on the X chromosome, genes located on the X chromosome are also silenced. Thus, N^6 -methyladenine developed a new role in epigenetic silencing in mammalian evolution distinct from its role in gene activation in other organisms. Our results demonstrate that N^6 -methyladenine constitutes a crucial component of the epigenetic regulation repertoire in mammalian genomes.

DNA methylation is a crucial component of epigenetic regulation that controls many important aspects of mammalian biology, such as imprinting, X chromosome inactivation and tumorigenesis^{1,2}. The prevailing dogma states that DNA methylation exclusively occurs on the fifth position of cytosine (5mC) in mammals, whereas the other modifications are absent, such as N^6 -methyladenine (N6-mA) which is predominantly present in prokaryotes and a limited number of eukaryotes³. Several reports have very recently expanded the list of organisms with N6-mA to three additional eukaryotes: insects (*Drosophila melanogaster*)⁴, nematodes (*Caenorhabditis elegans*)⁵ and green algae (*Chlamydomonas reinhardtii*)⁶; and intriguingly, these studies implicated N6-mA in gene activation^{4–6}, instead of repression, as is the case for 5mC repression. Despite this progress, the central issue regarding additional DNA modifications in mammals remained unresolved. A single report in the 1980s showed indirect evidence of N6-mA in mammalian genomes⁷; subsequent studies, however, were unable to confirm the presence of N6-mA in mammalian genomes⁸. A major function of 5mC in mammals is to control retrotransposons, for example, the long interspersed element 1 (LINE-1 or L1), a non-LTR family retrotransposon^{9,10}. Although the majority of the LINE-1 transposons, which have lost the 5' UTR and other regions proximal to the 5' end, are transcriptionally inactive¹⁰, several thousands of full-length (6–7 kb), young LINE-1 transposons (that emerged in the mouse genome less than 1.5 million years ago^{11,12}), which contain their own promoters at the 5' UTR, can be autonomously transcribed.

Incorporation of histone variant proteins, which carry significantly different primary sequences from the major histone isoforms, is another important aspect of epigenetic regulation¹³. These variants, which usually account for a very small fraction of the total histone pool, are deposited in critical genomic regions and play important roles in cell fate decisions and development¹³. It has been shown that the local

structure of histone variant-containing nucleosomes may be different from the canonical ones, consistent with the significant differences in protein (histone) primary sequences¹⁴. By the same token, it is conceivable that the altered nucleosome structures may be employed in accommodating variations in DNA structures, such as chemical modifications. In this work, we developed a single molecular real-time sequencing of chromatin immunoprecipitation-enriched DNA (SMRT-ChIP) approach to interrogate DNA modifications enriched at histone variant H2A.X deposition regions in mouse embryonic stem cells, leading to the identification of N6-mA in mouse embryonic stem cells and the associated demethylase, as well as revealing a novel evolved function of gene repression.

Identification of N6-mA in mouse embryonic stem cells

As SMRT sequencing usually requires high sequencing coverage to identify modified DNA bases^{15,16}, it is difficult to interrogate large mammalian genomes (2.8 Gb of *Mus musculus*, for example) with this approach¹⁶. Therefore we developed a SMRT-ChIP approach to interrogate specific genomic regions of interest (Fig. 1a). As H2A.X deposition is strongly associated with cell fate transitions in mammals¹⁷, we focused on H2A.X deposition regions in embryonic stem (ES) cells in the current study. DNA molecules residing in H2A.X deposition regions in mouse ES cells were subject to SMRT sequencing directly without PCR amplification (Methods). In total, 90% of SMRT-ChIP reads overlapped with H2A.X deposition regions identified by traditional ChIP-seq in a previous work¹⁷ (Extended Data Fig. 1a).

This approach identified N6-mA sites in H2A.X deposition regions with high confidence (398 sites at sequence coverage >30×, QV score ≥ 30 to 1,108 sites at sequence coverage >25×, QV score ≥ 20; see Extended Data Fig. 1b). A representative N6-mA site is shown in Fig. 1b. Several specific DNA motifs, which are different from H2A.X

¹Department of Genetics and Yale Stem Cell Center, Yale School of Medicine, New Haven, Connecticut 06520, USA. ²Pacific Biosciences, 1380 Willow Road, Menlo Park, California 94025, USA.

³Environmental Sciences & Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁴Department of Genetics and Genomic Sciences and Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York 10029, USA. ⁵Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA. ⁶Yale Stem Cell Center and Department of Cell Biology, Yale School of Medicine, New Haven, Connecticut 06520, USA. ⁷Department of Molecular Biophysics & Biochemistry, Yale Center for Genome Analysis, Yale School of Medicine, New Haven, Connecticut 06520, USA.

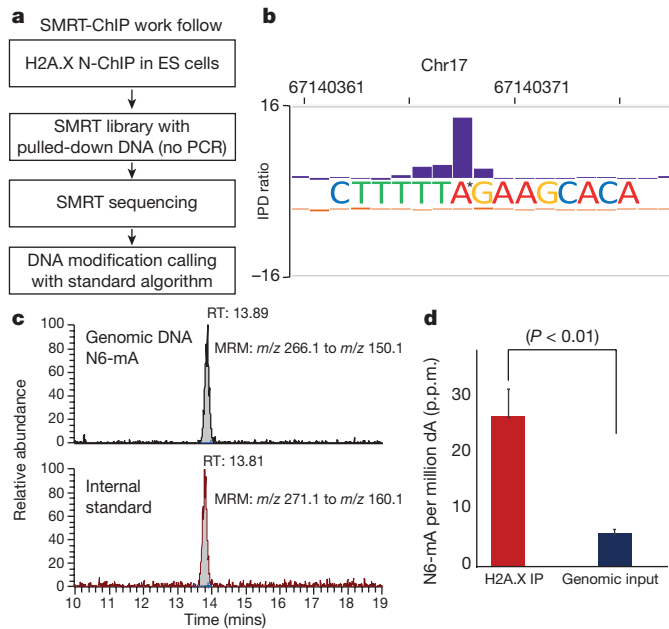


Figure 1 | A SMRT-ChIP approach identified N6-mA in mammalian genomes. **a**, Schematic of SMRT-ChIP. **b**, Sequencing tracks of N6-mA in ES cells. IPD ratio, inter-pulse distance ratio. **c**, Top: LC-mass spectrometry analysis of N6-mA ($m/z = 266.1$ to $m/z = 150.1$). Bottom: stable isotope labelled N6-mA ($m/z = 271.1$ to $m/z = 155.1$), internal standard. MRM, multiple reaction monitoring. **d**, Quantification of the LC-MS/MS results. $P < 0.01$, t -test; error bars, \pm s.e.m. of three biological replicates.

deposition motifs (Extended Data Fig. 1c), were significantly associated with these putative N6-mA sites, indicating that its distribution in the genome is controlled by yet unknown factors or pathways (Extended Data Fig. 1c). These N6-mA sites are enriched at intergenic, but not gene-rich regions ($P < 2.2 \times 10^{-16}$, Extended Data Fig. 1d).

We next confirmed the presence of N6-mA with mass spectrometry (MS). DNA molecules from the whole genome or H2A.X-deposition regions were subjected to an established and highly sensitive (LOQ: 1.6 fmol) mass spectrometry (liquid chromatography–mass spectrometry (LC-MS/MS)) approach¹⁸, which leverages stable isotope-labelled [¹⁵N₅] N6-mA as an internal standard for sample enrichment and quantification (Fig. 1c, d and Extended Data Fig. 2a). This approach identified N6-mA in embryonic stem cells (Fig. 1c); and resulted in an estimate of a frequency of 25–30 p.p.m. of deoxy-adenine (dA) in the H2A.X deposition regions for the N6-mA modification (Fig. 1d), a fourfold enrichment over the whole genomic input DNA samples (6–7 p.p.m.). We also investigated and found very low levels of N6-mA in other differentiated mouse cells and adult tissues (Extended Data Fig. 2b).

Importantly, none of the other known alkylation adducts, such as 1-methyladenine (N1mA), 3-methyladenine (N3mA) or 3-methylcytosine (N3mC)¹⁹, were detected from the H2A.X deposition region or whole genomic DNA samples (Extended Data Fig. 2c). Although it was reported that N1mA shares similar kinetic profiles to N6-mA in SMRT sequencing²⁰, our mass spectrometry approach which can distinguish N6-mA from N1mA, which ruled out this possible explanation of the SMRT-ChIP data (Extended Data Fig. 2d, e).

Alkbh1 encodes a demethylase for N6-mA in ES cells

We next focused on identifying the N6-mA demethylase. The mammalian *Alkbh* family genes, which contain the conserved Fe²⁺ ion and 2-oxo-glutarate-dependent, dioxygenase domain, were promising candidates²¹. Among these genes, the proteins encoded by *Alkbh2* and *Alkbh3* can efficiently remove 1mA or 3mC from DNA or RNA, but not N6-mA (see refs 19 and 21). *Alkbh1* is arguably the most intriguing member in this gene family: it shares the strongest similarity to bacteria demethylase *Alkb*, and yet only has negligible demethylation

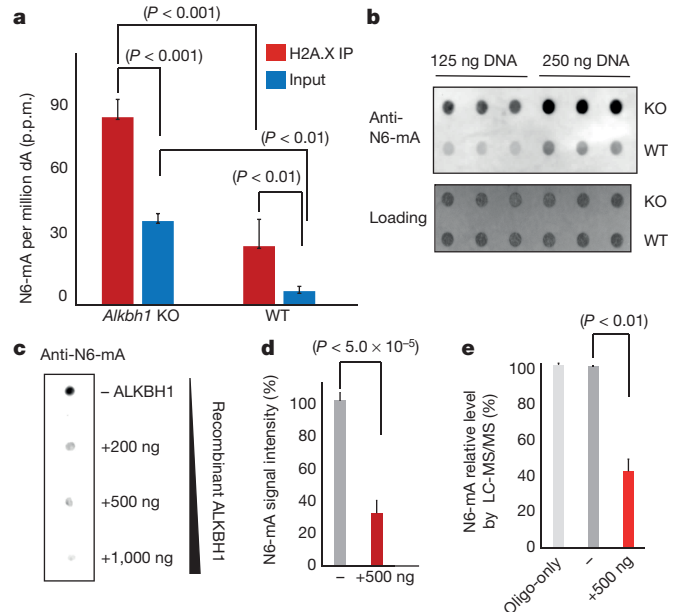


Figure 2 | *Alkbh1* is a demethylase for N6-mA in ES cells. **a**, Mass spectrometry analysis of N6-mA in *Alkbh1* knockout (KO) ES cells (P value determined by t -tests). **b**, Dot blotting of N6-mA in *Alkbh1* knockout or wild-type (WT) ES cells (in triplicates). **c**, *In vitro* demethylation reaction with recombinant ALKBH1 proteins monitored by dot blotting (Methods). **d**, Quantification of demethylation activity in three independent demethylase assays in **c** (P value $< 5.0 \times 10^{-5}$, t -test). **e**, *In vitro* demethylation reaction monitored by mass spectrometry (P value < 0.01 , t -test). Error bars, s.d. for three biological replicates.

activities on 3mC in comparison to *Alkbh2* and *Alkbh3* (see refs 19, 21). Additionally, an *Alkbh1* deficiency in mice results in 80% reduction of the litter size due to embryonic lethality among other phenotypes, indicating that *Alkbh1* plays a critical role in early development^{22,23}.

We generated *Alkbh1* homozygous knockout embryonic stem cell lines (referred to as *Alkbh1* knockout embryonic stem cells hereafter) via CRISPR/Cas9 technology (Extended Data Fig. 3a). Mass spectrometry analysis demonstrated that N6-mA levels in whole genomic input DNA or H2A.X deposition regions were both significantly increased (threefold to fourfold) in multiple *Alkbh1* knockout embryonic stem cell clones (Fig. 2a). Similar elevated N6-mA levels in *Alkbh1* knockout embryonic stem cells were confirmed by immunoblotting experiments with specific antibodies against N6-mA (Fig. 2b and Extended Data Fig. 3b–d). Previous work suggested that *Alkbh1* may regulate histone H2A K118 or K119 methylation in embryonic stem cells²⁴. We investigated and ruled out the possibility of *Alkbh1* being a histone demethylase, as H2AK118/119 is predominately non-methylated in wild-type or *Alkbh1* knockout ES cells (Extended Data Fig. 3e).

We investigated the catalytic activities of recombinant ALKBH1 proteins with *in vitro* demethylation assays. The recombinant ALKBH1 proteins were generated with $>95\%$ purity (Extended Data Fig. 3f). Recombinant ALKBH1 can efficiently reduce N6-mA level from single-stranded synthetic oligonucleotide substrates (Fig. 2c–e), while its activities towards dual- or hemi-methylated double-stranded substrates are much reduced, suggesting the demethylation may be coupled with transcription and/or replication *in vivo* (Extended Data Fig. 3g). Furthermore, these activities are dependent on Fe²⁺ ion and 2-oxoglutarate, as expected for an active dioxygenase (Extended Data Fig. 3h).

The catalytic activities of ALKBH1 were further substantiated by a point mutant at a critical residue (D233A) that may coordinate the Fe²⁺ ion. Corroborated by the much reduced activities of the recombinant mutant proteins (D233A) (Extended Data Fig. 3i, j), the increase of N6-mA in *Alkbh1* knockout mouse ES cells could be efficiently rescued by ectopic expression of wild-type but not mutant *Alkbh1* (Extended Data Fig. 3k, l).

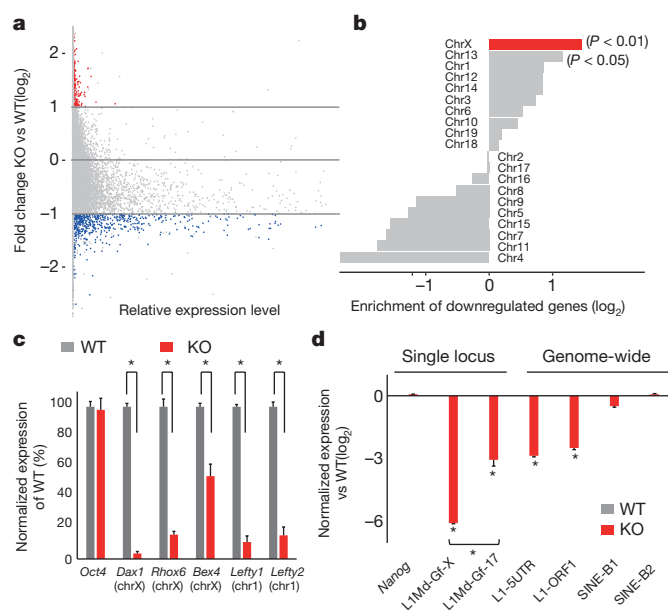


Figure 3 | *Alkbh1* deficiency silences genes on the X chromosome and young full-length L1 elements. **a**, RNA-seq analysis of *Alkbh1* knockout ES cells versus wild-type controls. Blue: most highly downregulated genes, red: upregulated genes (false positives, see main text). **b**, Downregulated genes were most enriched on X chromosome ($P < 0.01$, binomial test) and Chr13 to a lesser extent ($P < 0.05$, binomial test). **c**, qRT-PCR analysis of downregulated genes ($*P < 0.05$, t -test). **d**, RT-qPCR of transposon expression ($*P < 0.01$, t -test). L1Md-Gf-X: a young full-length L1 on Chr-X. L1Md-Gf-17: a young full-length L1 on Chr17. Error bars, \pm s.e.m. of three technical replicates.

N6-mA suppresses transcription on ChrX

The identification of *Alkbh1* as a N6-mA demethylase enabled us to test the functions of N6-mA in ES cells. As this modification may be an important component of epigenetic regulation of gene expression, we used a RNA-seq approach to interrogate the transcriptome of *Alkbh1* knockout ES cells. Our analysis demonstrated that 550 genes were significantly downregulated (fragments per kilobase of transcript per million mapped reads (FPKM) > 5 , false discovery rate (FDR) < 0.05 , fold change > 2 or < 0.5 , from Cuffdiff2) (Fig. 3a, and Supplementary Table 1), which can be verified by the RT-qPCR approach (Extended Data Fig. 4a). Although a small number of genes with low expression levels (70) were initially identified as upregulated by the RNA-seq analysis, they were probably false positives which cannot be verified with an RT-qPCR approach (0/5, Extended Data Fig. 4a, b), indicating that increasing the N6-mA level in ES cells leads to gene silencing. Gene ontology analysis showed that the most highly downregulated genes are enriched for developmental factors or lineage specifying genes (Extended Data Fig. 4c). On the other hand, the expressions of pluripotency genes, such as *Oct4* and *Nanog*, were unaltered and *Alkbh1* knockout ES cells maintained the undifferentiated morphology and were able to self-renew.

Unexpectedly, the genomic locations of the downregulated genes have a strong chromosome bias ($P < 0.01$, binomial test): they are most significantly enriched on the X chromosome, whereas modestly enriched on Chr13 ($P < 0.05$, binomial test), but not on the other chromosomes (Fig. 3b). qRT-PCR analysis confirmed the downregulation of the X chromosome genes, together with other genes on autosomes (Fig. 3c). These results indicate that accumulation of N6-mA represses transcription on the X chromosome.

To test this hypothesis, we investigated the expression of young full-length LINE-1 transposons (L1 elements) which are specifically enriched on the X chromosome (see Fig. 4 and refs 25, 26). Owing to their unique sequences, the expression of such L1 elements can be interrogated and distinguished from other L1 subfamilies²⁷. Our results

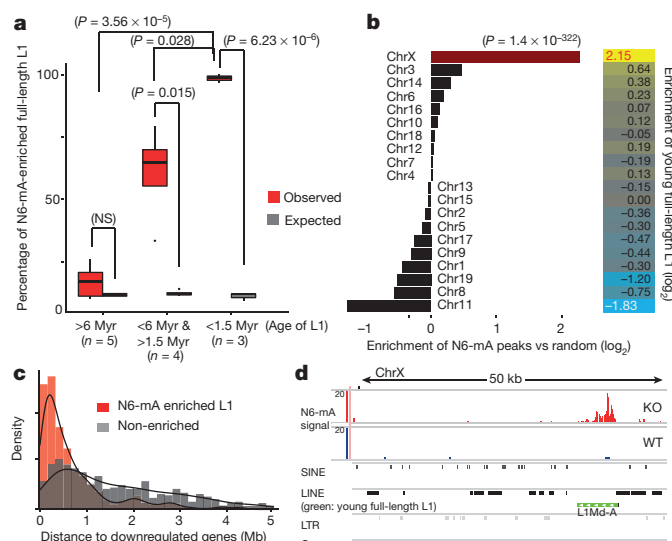


Figure 4 | N6-mA is enriched at young full-length L1 elements, which are located in the vicinity of the downregulated genes in *Alkbh1* knockout ES cells. **a**, Enrichment of N6-mA on full-length L1 elements (P value determined by t -test). **b**, Left: relative enrichment of N6-mA peaks on each chromosome ($P = 1.4 \times 10^{-322}$, binomial test). Right: relative enrichment of young full-length L1s on each chromosome. **c**, Normalized frequency of full-length L1 elements was plotted as a function of their genomic distance to downregulated genes (red, N6-mA enriched, median: 424 kb; grey, non-enriched, median: 1.6 Mb). **d**, The *Dax1* gene locus.

demonstrated that a young full-length L1 (belong to the L1Md-Gf subfamily^{11,12}) located on the X chromosome is more highly repressed (more than 60-fold) than its counterpart located on Chr17 (Fig. 3d). These results indicated that the L1 density may affect the silencing effects of N6-mA. A qRT-PCR approach targeting the 5' UTR or open reading frame 1 (ORF1), which are usually retained in young full-length L1 elements, but not old truncated L1 elements¹⁰, also demonstrated a significant decrease of L1 expression, whereas the SINE family transposons were almost unaffected (Fig. 3d). Additionally, analyses of the transposons transcripts in the RNA-seq experiments confirmed the downregulation of the young full-length L1 subfamilies (Methods and Extended Data Fig. 4d). These results raised the intriguing possibility that genes and young full-length L1 elements on X chromosomes may be co-regulated by N6-mA.

N6-mA specifically targets young full-length L1 elements

The results suggest that N6-mA adopts a new function of transcriptional silencing in mammals, whereas it is implicated in gene activation in other species³⁻⁶. To further investigate N6-mA function, we sought to identify the differentially methylated regions (DMR) of N6-mA in *Alkbh1* knockout ES cells.

As there is a global increase of N6-mA in *Alkbh1* knockout cells as indicated by mass spectrometry analyses (Fig. 2). The SMRT-ChIP approach can only interrogate H2A.X deposition regions (Fig. 1a), so we performed a N6-mA DIP-seq (N6-mA DNA immunoprecipitation with anti-N6-methyladenine antibodies followed by next-generation sequencing) experiment (Methods). To validate this approach, we first investigated and determined its detection limit and lineage response range by a 'spike-in experiment' (Methods). With this approach, the detection limit is around 10–15 p.p.m. N6-mA (of adenine), while this approach cannot distinguish N6-mA from unmodified adenines at 5 p.p.m. N6-mA levels in *Alkbh1* knockout cells (30–35 p.p.m.) is within the lineage range of this approach (20 to 120 p.p.m.) (Extended Data Fig. 5a).

Consistent with the genome-wide upregulation, N6-mA DIP-seq identified 37,581 N6-mA sites in *Alkbh1* knockout ES cells, in agreement with the estimate (35,000–40,000 sites) based on mass spectrometry results (30–35 p.p.m.). On the other hand, the N6-mA peaks in wild-type

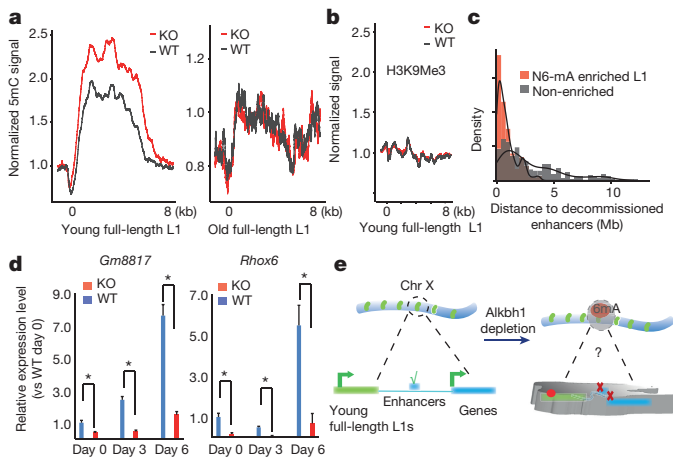


Figure 5 | N6-mA upregulation induced transcriptional silencing on the X chromosome, which is persistent during differentiation. **a**, Aggregation of 5mC. **b**, Aggregation of H3K9Me3 signals. **c**, Normalized frequency of decommissioned enhancers was plotted as a function of their genomic distance to full-length L1 elements red, N6-mA enriched, median: 484 kb; grey, non-enriched, median: 2 Mb. **d**, RT-qPCR analysis of the *Gm8817* and *Rhox6* genes (on the X chromosome) during embryoid body differentiation. * $P < 0.05$, *t*-test; error bars, \pm s.e.m. of three biological replicates. **e**, Schematics of *Alkbh1* and N6-mA functions (see main text).

ES cells are under-represented as N6-mA frequency is only 6–7 p.p.m. in these cells. We also used SMRT-ChIP approach to interrogate N6-mA distribution in H2A.X deposition regions in *Alkbh1* knockout ES cells (Extended Data Fig. 5b, c). Our results demonstrated that putative N6-mA sites called by SMRT-ChIP at various cutoffs (sequences coverage: $10\times$ to $30\times$; QV: 20–30) significantly ($P < 1.0 \times 10^{-5}$; observed versus permutation) overlap with those identified by DIP-seq. In addition, the percentage of overlap increases with rising sequencing coverage and QV scores. These results further validate the SMRT-ChIP approach.

N6-mA peaks called from DIP-seq are enriched in intergenic regions, but not gene-coding regions (Extended Data Fig. 5e). Further analysis showed that N6-mA are deposited at LINE elements (Extended Data Fig. 5f), especially full-length L1 elements, but not the truncated ones (Extended Data Fig. 5g). Remarkably, N6-mA deposition at L1 elements is inversely correlated with their evolutionary age; over 99% of the young full-length L1 elements are enriched for N6-mA, whereas no such enrichment is observed on old L1 elements (Fig. 4a and Extended Data Fig. 5h). One of the major differences between the young and old L1 elements is that the former retain the 5' UTR and ORF1 regions, whereas old L1 elements gradually lost their 5' UTR and ORF1 during multiple rounds of remobilization in evolution and therefore became inactive¹⁰. N6-mA deposition is biased at the 5' UTR and ORF1 regions rather than at the 3' UTR (Extended Data Fig. 6a). This enrichment pattern was confirmed using a qPCR approach (Extended Data Fig. 6b).

Young full-length L1 elements are strongly enriched on X chromosomes over autosomes^{25,26}, and our analysis corroborated this longstanding observation (Fig. 4b, $P = 1.4 \times 10^{-322}$). Consistent with this, N6-mA peaks in *Alkbh1* knockout ES cells are also significantly enriched on the X chromosome over autosomes (Fig. 4b, $P = 1.4 \times 10^{-322}$). Therefore, these results are consistent with the downregulation of young full-length LINE-1 sequences and protein-coding genes located on X chromosomes (Fig. 3).

In classic epigenetic silencing pathways, the distance between the silencing centre and genes is a critical determinant of silencing. Consistent with notion, further analysis showed that the downregulated genes are located much closer to the N6-mA-enriched L1 elements (median: 424 kb) than to the non-enriched ones (median: 1.6 Mb) (Fig. 4c). Furthermore, the distances from downregulated genes to the N6-mA-enriched L1 elements fall within a narrow range (25–75%: 196–925 kb), while such distances to the non-enriched ones display greater variations (688 kb to 3.2 Mb,

Extended Data Fig. 7a). The *Nr0b1* (also known as *Dax1*) gene that was significantly downregulated in *Alkbh1* knockout ES cells (Fig. 3) was not enriched for N6-mA; it is, however, located 30 kb from a N6-mA-enriched young full-length L1 (Fig. 4d, green). Other transposons located in this genomic region are not enriched for N6-mA (Fig. 4d).

The distances between either the ES-cell expressing genes in wild-type ES cells (FPKM > 5.0 in RNA-seq) or downregulated genes in *Alkbh1* knockout ES cells and young full-length L1 elements on Chr13 are significantly shorter than the other autosomes ($P < 2.2 \times 10^{-16}$, Extended Data Fig. 7b, c). However, on a few chromosomes which are devoid of the downregulated genes in *Alkbh1* knockout ES cells, especially Chr11 and Chr4 (see Fig. 3), such distances are significantly longer than the other chromosomes (L1 to ES-cell-expressing genes: around 1,000 kb, $P < 2.98 \times 10^{-13}$; L1 to downregulated genes: around 800 kb, $P \leq 0.01$ Extended Data Fig. 7b, c).

Increasing N6-mA levels leads to silencing

Our results indicated that N6-mA may have a direct effect on the transcription of L1 elements and their neighbouring genes. Thus, we investigated the impacts of N6-mA deposition on young full-length L1 elements and their neighbouring genes by interrogating the genome-wide deposition of several key epigenetic marks implicated in transcriptional regulation.

First, we focused on the effects of N6-mA on young full-length L1 elements. Our analysis demonstrated that although the genome-wide distribution and intensities of 5mC methylation sites are similar in *Alkbh1* knockout and the wild-type control (Extended Data Fig. 8a), the 5mC level on the young full-length L1 elements is modestly higher in *Alkbh1* knockout than wild-type control, while such differences are not observed on old L1 elements (Fig. 5a) or SINEs (Extended Data Fig. 8b). Other epigenetic silencing marks, such as H3K9me3 (Fig. 5b), H3K27me3 and H2A.X, are deposited on young full-length L1 elements at similar levels (Extended Data Fig. 8). Although these results are consistent with previous works showing that the young L1 elements are silenced by 5mC in human ES cells¹², additional mechanisms may be also involved as the effects of 5mC seem to be modest.

We interrogated the epigenetic status of the enhancers and the results demonstrated that 450 enhancers (Supplementary Table 3) are decommissioned, as their H3K27Ac levels are significantly decreased in *Alkbh1* knockout ES cells (one locus shown in Extended Data Fig. 8c). These decommissioned enhancers are located much closer to N6-mA-enriched L1 elements (median: 485 kb) than non-enriched ones (2.03 Mb, Fig. 5c). Furthermore, such distances fall into a much narrower range (25–75%: 197–985 kb) than those to the non-enriched ones (806 kb to 3.8 Mb) (Extended Data Fig. 8d). Furthermore, the H3K4Me3 levels are reduced at the transcription start sites of the downregulated genes (but not at the unaffected ones) (Extended Data Fig. 8e). These data demonstrate that N6-mA deposition at L1 is correlated with the downregulation of nearby genes at the transcription level.

The potential effects of N6-mA deposition on X chromosome genes during differentiation was investigated. Embryoid body formation and differentiation assays were performed. Although the *Alkbh1* knockout ES cells are able to differentiate, the cell fate decisions (relative gene expression levels of the three germ layer marks after differentiation) are imbalanced, as is consistent with previous reports (Extended Data Fig. 9). X chromosome genes, such as *Gm8817* and *Rhox6* (ref. 28), failed to be activated to the normal level in *Alkbh1* knockout ES-cell-derived embryoid bodies (Fig. 5d), indicating that N6-mA modifications have long-lasting effects on activation of the genes during differentiation.

Discussion

We have developed a novel approach (SMRT-ChIP) to interrogate DNA modifications in specific genomic regions, resulting in the discovery of N6-mA in the mammalian genome and the identification of the demethylase *Alkbh1*. These findings challenge the prevailing paradigm that 5mC is the only form of DNA methylation in the mammalian genome.

N6-mA seems to have adopted new functions during evolution. In mammalian ES cells, N6-mA accumulation on young full-length L1 elements correlates with direct silencing of such L1 elements, as well as decommissioning of nearby enhancers and genes, which is in direct contrast to the role of N6-mA in simple eukaryotes and invertebrates^{4–6}. In addition, the only Fe²⁺, 2KG-dependent dioxygenase orthologue in the *Drosophila* genome has been reported to demethylate N6-mA in DNA⁴ and oxidize 5mC in RNA²⁹, whereas the functions of mammalian orthologues (*Tet1–3* and *Alkbh1–8* genes) are much divergent. N6-mA silencing of L1 transposons in *Alkbh1*-deficient cells is inversely correlated with the evolutionary age of the transposon; the full-length, young L1 elements are specifically targeted and silenced by N6-mA. Although the precise reasons for this remains elusive, our results showed that N6-mA deposition is strong on the unique 5' UTR and ORF1 regions of such L1 elements which harbour the promoters. These results also suggest that *Alkbh1* must be targeted to these regions in wild-type ES cells and future investigation will determine molecular underpinning of this specific targeting. Furthermore, as young full-length L1 elements are strongly enriched on the X chromosome, N6-mA deposition displays a strong bias towards the X chromosome. As such, our findings herein may shed new light on the longstanding hypothesis of L1 function during X inactivation proposed by M. Lyon³⁰. Although young full-length L1 elements are active during early embryogenesis³¹, constant activation may cause genomic instability as they are capable of reintegration^{10,11}, which implies the existence of a previously unknown silencing mechanism. We favour the view that N6-mA-mediated silencing plays an important role in safeguarding active L1 elements in mammalian genomes. The levels of N6-mA are controlled precisely by *Alkbh1* in ES cells such that they favour L1 transcription while preventing it from succumbing to overactivation and genomic instability, which is reminiscent of the function of a rheostat (Fig. 5e). In addition, LINE-1s are inactive in a group of South American rodents, in which a new family of endogenous retrotransposons (mysTR) has emerged³². It will be interesting to determine the presence and functions of N6-mA in these rodents. During the review process of this manuscript, Koziol *et al.* reported the presence of N6-mA in adult mouse tissues³³. However, N6-mA levels in these tissues seem to be lower than the detection limit of the DIP-seq approach³³. Note that different statistical thresholds were applied in their bioinformatic analyses³³ and discrepancies between the two studies still need to be resolved. Taken together, the discovery of N6-mA in mammalian ES cells sheds new light on epigenetic regulation during early embryogenesis and may have impacts in the fields of epigenetics, stem cells and developmental biology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 August 2015; accepted 7 March 2016.

Published online 30 March; corrected online 20 April 2016

(see full-text HTML version for details).

- Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14**, 204–220 (2013).
- Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
- Heyn, H. & Esteller, M. An adenine code for DNA: a second life for N⁶-methyladenine. *Cell* **161**, 710–713 (2015).
- Zhang, G. *et al.* N⁶-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906 (2015).
- Greer, E. L. *et al.* DNA methylation on N⁶-adenine in *C. elegans*. *Cell* **161**, 868–878 (2015).
- Fu, Y. *et al.* N⁶-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**, 879–892 (2015).
- Achwal, C. W., Iyer, C. A. & Chandra, H. S. Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, *Drosophila* and mealybug DNA. *FEBS Lett.* **158**, 353–358 (1983).
- Ratel, D. *et al.* Undetectable levels of N6-methyl adenine in mouse DNA: Cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. *FEBS Lett.* **580**, 3179–3184 (2006).
- Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).

- Goodier, J. L. & Kazazian, H. H. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135**, 23–35 (2008).
- Goodier, J. L., Ostertag, E. M., Du, K. & Kazazian, H. H. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.* **11**, 1677–1685 (2001).
- Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
- Banaszynski, L. A., Allis, C. D. & Lewis, P. W. Histone variants in metazoan development. *Dev. Cell* **19**, 662–674 (2010).
- Jin, C. & Felsenfeld, G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* **21**, 1519–1529 (2007).
- Fang, G. *et al.* Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nature Biotechnol.* **30**, 1232–1239 (2012).
- Davis, B. M., Chao, M. C. & Waldor, M. K. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Curr. Opin. Microbiol.* **16**, 192–198 (2013).
- Wu, T. *et al.* Histone variant H2A.X deposition pattern serves as a functional epigenetic mark for distinguishing the developmental potentials of iPSCs. *Cell Stem Cell* **15**, 281–294 (2014).
- Lu, K., Collins, L. B., Ru, H., Bermudez, E. & Swenberg, J. A. Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. *Toxicol. Sci.* **116**, 441–451 (2010).
- Sedgwick, B. Repairing DNA-methylation damage. *Nature Rev. Mol. Cell Biol.* **5**, 148–157 (2004).
- Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**, 461–465 (2010).
- Shen, L., Song, C.-X., He, C. & Zhang, Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.* **83**, 585–614 (2014).
- Müller, T. A., Yu, K., Hausinger, R. P. & Meek, K. ALKBH1 is dispensable for abasic site cleavage during base excision repair and class switch recombination. *PLoS ONE* **8**, e67403 (2013).
- Nordstrand, L. M. *et al.* Mice lacking *Alkbh1* display sex-ratio distortion and unilateral eye defects. *PLoS ONE* **5**, e13827 (2010).
- Ougland, R. *et al.* ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. *Stem Cells* **30**, 2672–2682 (2012).
- Abrusán, G., Giordano, J. & Warburton, P. E. Analysis of transposon interruptions suggests selection for L1 elements on the X chromosome. *PLoS Genet.* **4**, e1000172 (2008).
- Bailey, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl Acad. Sci. USA* **97**, 6634–6639 (2000).
- Chow, J. C. *et al.* LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**, 956–969 (2010).
- Liu, C., Tsai, P., García, A.-M., Logeman, B. & Tanaka, T. S. A possible role of *Reproductive Homeobox 6* in primordial germ cell differentiation. *Int. J. Dev. Biol.* **55**, 909–916 (2011).
- Delatte, B. *et al.* Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282–285 (2016).
- Lyon, M. F. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* **80**, 133–137 (1998).
- Fadloun, A. *et al.* Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nature Struct. Mol. Biol.* **20**, 332–338 (2013).
- Erickson, I. K., Cantrell, M. A., Scott, L. & Wichman, H. A. Retrofitting the genome: L1 extinction follows endogenous retroviral expansion in a group of murid rodents. *J. Virol.* **85**, 12315–12323 (2011).
- Koziol, M. J. *et al.* Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nature Struct. Mol. Biol.* **23**, 24–30 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Z. Li, K. Hwang and A. Leung for critical reading of the manuscript and the members of the Xiao laboratory for critical discussion. Thanks to L. Geng for helping HiSeq2000 sequencing. This work is funded by R01GM114205-01 (A.X.). T.P.W. is partially supported by CT Stem Cell Foundation (11SCA34). The Fang lab is partially supported by R01 GM114472-01 (G.F.). Mass spectrometry was supported by R01GM106024, S10OD018445 and P20GM103429. The UNC Mass Spectrometry Facility Core was supported by the National Institutes of Environmental Health Sciences (NIEHS) Superfund Basic Research Program (P42 ES005948), and NIEHS Center for Environmental Health and Susceptibility (P30 ES010126).

Author Contributions A.X. conceived the hypothesis, designed the study and wrote the paper, and provided support and guidance for this work; T.P.W. designed and performed the majority of the experiments, analysed the genomic data, generated figures and interpreted the results; T.W. characterized the *Alkbh1* mutant and performed demethylation assays; K.L. helped with bioinformatics analysis. Y.L. provided technical help. L.H., M.S., S.Z. and G.F. assisted on SMRT sequencing and data analysis. Y.L. and J.A.S. performed the mass spectrometry analysis of N6-mA. S.D.B., S.G.M. and A.J.T. performed MS analysis on histone methylation and recombinant ALKBH1 proteins.

Author Information All sequencing data were deposited in the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE71866. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.X. (Andrew.Xiao@Yale.edu).

METHODS

Mouse ES cell culture. Mouse TT2 ES cells were cultured on gelatin coating plates with recombinant LIF. ES cells were grown in DMEM supplemented with 15% fetal bovine serum, 1% non-essential amino acids, 2 mM L-glutamine, 1,000 units of mLIF (EMD Millipore), 0.1 mM β -mercaptoethanol (Sigma) and antibiotics.

Generation of *Alkbh1* knockout ES cell lines with CRISPR–Cas9. A doxycycline (Dox)-inducible Cas9–eGFP ES cell line was established with TT2 ESC. Guide RNA oligos (5′-accgAGTGCCTCTGGCATCCCGGG-3′, 5′-aacCCCGGGATGC CAGAGGCACT-3′) were annealed and cloned into a pLKO.1-based construct (Addgene: 52628). Guide RNA virus was made in 293FT cells and infected inducible Cas9 ES cells. ES cells were first selected with Puromycin (1 μ g ml^{−1}) for two days, and Dox (0.5 μ g ml^{−1}) was added to induce Cas9–eGFP expression for 24 h. ES cells were then seeded at low density to obtain single-derived colonies. Then, 72 ES cell colonies were randomly picked up and screened by PCR–enzyme digestion that is illustrated in Extended Data Fig. 3a. PCR screening primers flanking guide RNA sequence were designed as following: 5′-AGGCAGATTCTGAGTTCAAGG-3′ and 5′-TTTAGTCATGTGCTTGCCAGG-3′.

PCR products were digested by XmaI overnight at 37 degrees and separated on 2% agarose gel. A total of 8 mutants from which PCR products show resistance to XmaI digestion were subjected to DNA sequencing. Clones that harbour deletion and coding frame shift (premature termination mutation) were expanded and used in this study.

Expression of ALKBH1 protein in 293FT cells and generation of ALKBH1 mutation proteins. Human Alkbh–Flag DNA sequence was inserted into pCW lenti-virus based vector (puromycin or hygromycin resistance). The amino acid of D233 was mutated to A by QuickChange Site-Directed Mutagenesis (QuickChange II XL Site-Directed Mutagenesis Kit, number 200521, Agilent) according to the manual. For Alkbh1 rescue experiment, wild-type and D233A mutated Alkbh1 constructs were introduced to Alkbh1 knockout ES cells, pCW-Hygromycin was chosen as control. After infections, the cells were selected with hygromycin at 200 μ g ml^{−1} for 4 days, and then the cells were expanded to isolate genomic DNA for N6-mA dot blotting or other tests.

The 293FT cells were transfected with pCW-hAlkbh1 and pCW-hAlkbh1-D233A mutant plasmids along with package plasmids of pMD2.G and pSPAX2. Culture medium was changed 10 h after transfection. The viruses were collected and concentrated 24 and 48 h after transfection according to manufacturer's instructions (Lenti-X Concentrator, Clontech). To establish stable expression of hAlkbh1 and hAlkbh1-D233A cell lines, 293FT cells were infected the corresponding virus, and then select with puromycin at 1 μ g ml^{−1} for 4 days. The stable cell lines of hAlkbh1-293FT and D233A-293FT were expanded to purify the proteins according to the previous reported method with some modifications³⁴. Briefly, M2 Flag antibody was added to the nuclear extract and incubated overnight, and then Dynabeads M-280 (sheep anti-mouse IgG, from Life Technology) was added to the above solution and incubated for 3–4 h. Subsequently, the beads were separated from the solution and washed clean with washing buffer³⁴. Finally, the beads were eluted with 3 \times Flag peptides, followed by standard chromatography purification to 95% purity. Proteins were analysed by mass spectrometry.

ALKBH1 demethylase assays. Demethylation assays were performed in 50 μ l volume, which contained 50 pmol of DNA oligos and 500 ng recombinant ALKBH1 (or D233A mutant) protein. The reaction mixture also consisted of 50 μ M KCl, 1 mM MgCl₂, 50 μ M HEPES (pH = 7.0), 2 mM ascorbic acid, 1 mM-KG, and 1 mM (NH₄)₂Fe(SO₄)₂·6H₂O. Reactions were performed at 37 degrees for 1 h and then stopped with EDTA followed by heating at 95 degrees for 5 min. Then the reaction product was subjected to dot blotting. Substrate sequences are listed in Supplementary Table 2.

Dot blotting. First, DNA samples were denatured at 95 degrees for 5 min, cooled down on ice, neutralized with 10% vol of 6.6 M ammonium acetate. Samples were spotted on the membrane (Amersham Hybond-N+, GE) and air dry for 5 min, then UV-crosslink (2 \times auto-crosslink, 1800 UV Stratalinker, STRATAGENE). Membranes were blocked in blocking buffer (5% milk, 1% BSA, PBST) for 2 h at room temperature, incubated with 6mA antibodies (202-003, Synaptic Systems, 1:1000) overnight at 4 degrees. After 5 washes, membranes were incubated with HRP linked secondary anti-rabbit IgG antibody (1:5,000, Cell Signaling 7074S) for 30 min at room temperature. Signals were detected with ECL Plus Western Blotting Reagent Pack (GE Healthcare).

Single molecule real-time sequencing (SMRT) library construction of genomic DNA samples and PCR control. DNA samples were purified by standard N-ChIP protocol. 5 μ g anti-H2A.X antibodies were used per 10 million cells. DNA (250 ng) from ChIP pull-down were converted to SMRTbell templates using the PacBio RS DNA Template Preparation Kit 1.0 (PacBio catalogue number 100-259-100) following manufacturer's instructions. Control samples were amplified by PCR (18 cycles). In brief, samples were end-repaired and ligated to blunt adaptors.

Exonuclease incubation was carried out in order to remove all unligated adaptors. Samples were extracted twice (0.6 \times AMPure beads) and the final 'SMRTbells' were eluted in 10 μ l embryoid bodies. Final quantification was carried out on an Agilent 2100 Bioanalyzer with 1 μ l of library. The amount of primer and polymerase required for the binding reaction was determined using the SMRTbell concentration (ng μ l^{−1}) and insert size previously determined using the manufacturer-provided calculator. Primers were annealed and polymerase was bound using the DNA/Polymerase Binding Kit P4 (PacBio catalogue number 100-236-500) and sequenced using DNA sequencing reagent 2.0 (PacBio catalogue number 100-216-400). Sequencing was performed on PacBio RS II sequencer using SMRT Cell 8Pac V3 (PacBio catalogue number 100-171-800). In all sequencing runs, a 240 min movie was captured for each SMRT Cell loaded with a single binding complex.

Detection of modified nucleotides with SMRT sequencing data. Base modification was detected using SMRT Analysis 2.3.0 (Pacific Biosciences), which uses previously published methods for identifying modified bases based on inter-pulse duration ratios in the sequencing data³⁵. All calculations used the *Mus musculus* mm10 genome as a reference. For the detection of modified bases in individual samples, the RS_Modification_Detection.1 protocol was used with the default parameters. Modifications were only called if the computed modification QV was better than 20, corresponding to $P < 0.01$ (versus in silico model, Welch's t -test). The *in silico* model considers the IPDs from the eight nucleotides 5′ through the three nucleotides 3′ of the site in question. Only the sites with a sequencing coverage higher than 25 fold were used for subsequent analyses. To assess the significance of the overlap between N6-mA sites by SMRT-ChIP and peaks from DIP-seq, intersection with DIP-seq peaks was analysed for each of the N6-mA site called by SMRT-ChIP. To assess if the overlap is higher than expected by random chance, a permutation based approach was used, in which we randomly shuffle the original mapping between "As" that meet coverage cutoff and their corresponding QV scores, and estimated the expected overlap by random chance. As preparation for PacBio RS II sequencing, these relatively short DNA fragments (200–1,000 base pairs on average) were made topologically circular, allowing each base to be read many times by a single sequencing polymerase. Thus, the coverage requirement for modification detection was achieved both by sequencing different fragments pulled down from the same genomic regions and by sequencing the same fragment with many passes. Of note, the SMRT-ChIP approach did not identify more N6-mA sites in *Alkbh1* knockout cells than wild-type cells. Although the exact reason remain to be identified, our analysis showed that much fewer adenines are sequenced at a comparable coverage in *Alkbh1* knockout cells than wild-type cells (Extended Data Fig. 5c and Extended Data Fig. 1b), presumably due to the difficulty of using native ChIP approach to isolate H2A.X-deposition regions from *Alkbh1* knockout cells because of heterochromatinization.

N6-mA-DNA-IP sequencing and analysis. Genomic DNA from wild-type or knockout ES cells was purified with DNeasy kit (QIAGEN, 69504). For each sample, 5 μ g DNA was sonicated to 200–500 bp with Bioruptor. Then, adaptors were ligated to genomic DNA fragments following the Illumina protocol. The ligated DNA fragments were denatured at 95 degree for 5 min. Then, the single-stranded DNA fragments were immunoprecipitated with 6 mA antibodies (5 μ g for each reaction, 202-003, Synaptic Systems) overnight at 4 degrees. N6-Me-dA enriched DNA fragments were purified according to the Active Motif hMeDIP protocol. IP DNA and input DNA were PCR amplified with Illumina indexing primers. The same volume WT and KO DNA samples were subjected to multiplexed library construction and sequencing with Illumina HiSeq2000. After sequencing and filter, high quality raw reads were aligned to the mouse genome (UCSC, mm10) with bowtie (2.2.4, default)³⁶. By default, bowtie searches for multiple alignments and only reports the best match; for repeat sequences, such as transposons, bowtie reports the best matched locus or random one from the best-matched loci. After alignment, N6-mA enriched regions were called with SICER (version 1.1, FDR $< 1.0 \times 10^{-15}$, input DNA as control)³⁷. Higher FDR cut-off could not further reduce N6-mA peak number. MACS2 was also used for peak calling, which generated similar results as SICER. Part of the data analysis was done by in-house customized scripts in R, Python or Perl. Genomic DNA samples from mouse fibroblast cells (where the endogenous N6-mA level is undetectable) were spiked with increasing amount of N6-mA-containing, or unmodified (control), oligonucleotides, and the N6-mA levels were determined by qPCR approach after DIP and library construction.

5mC DNA-IP sequencing. Followed manufacture's protocol (Active Motif 5mC MeDIP kit). The 5 mC data processed with MEDIPS in Bioconductor, and in-house scripts in R, Python or Perl.

ChIP-sequencing and data analysis pipeline. Native chromatin immunoprecipitation (N-ChIP) assay was performed as previously described. 10 million ES cells were used for each ChIP and massive parallel sequencing (ChIP-seq) experiment. Cell fractionation and chromatin pellet isolation were performed as

described. Chromatin pellets were briefly digested with micrococcal nuclease (New England BioLabs) and the mononucleosomes were monitored by electrophoresis. Co-purified DNA molecules were isolated and quantified (100–200 ng for sequencing). Co-purified DNA and whole cell extraction (WCE) input genomic DNA were subject to library construction, cluster generation and next-generation sequencing (Illumina HiSeq 2000).

The output sequencing reads were filtered and pre-analyzed with Illumina standard workflow. After filtration, the qualified tags (in fastq format) were aligned to the mouse genome (UCSC, mm10) with bowtie (2.2.4, default)³⁶. Then, these aligned reads were used for peak calling with the SICER algorithm (input control was used as control in peak calling).

Bioinformatics analysis of epigenetics ChIP Sequencing data. H3K4Me1 and H3K27Ac ChIP-seq data were aligned to mouse genome (mm10) and peaks were called with SICER. H3K4Me1 and H3K27Ac enriched regions were defined as enhancers. Then, RSEG³⁸ (mode 3) was to call the H3K27Ac differentiated regions. Decommissioned enhancers in KO cells are determined by H3K27Ac downregulation (compared to wild-type cells).

Detection of H3K4Me3 in knock-out cells with ChIP-qPCR. Native ChIP-qPCR assay was used to validate H4K4Me3 at levels on gene promoters (Extended Data Fig. 8). All procedures were similar to what has been described in ChIP-seq experiments, except that the co-purified DNA molecules were diluted and subject to qPCR (histone H3K4Me3 antibodies: Abcam Ab8580). Real-time PCR was performed with SybrGreen Reagent (Qiagen, QuantiTect SYBR Green PCR Kit, Cat: 204143) and quantified by a CFX96 system (BioRAD, Inc.).

RNA-seq and confirmation by RT-qPCR approaches. RNA was extracted with miRNeasy kit (QIAGEN, 217004) and standard RNA protocol. The quality of RNA samples was measured using the Agilent Bioanalyzer. Then, RNA was prepared for sequencing using standard Illumina 'TruSeq' single-end stranded or 'Pair-End' mRNA-seq library preparation protocols. 50 bp of single-end and 100 bp of pair-end sequencing were performed on an Illumina HiSeq 2000 instrument at Yale Stem Cell Center Genomics Core. RNA-seq reads were aligned to mm9 with splicing sites library with Tophat³⁹ (2.0.4, default parameters). The gene model and FPKM were obtained from Cufflink2. The differentially expressed genes were identified by Cuffdiff⁴⁰ (2.0.0, default parameters). To make sure the normalization is appropriate, the data were also analysed with DESeq2 (default parameters), which generated similar results (Extended Data Fig. 4b). For transposons analysis, unique best alignment reads were used (alignment with bowtie (0.12.9), -m 1; or BWA) and calculated RPKM for each subfamily. For qPCR, the cDNA libraries were generated with First-strand synthesis kit (Invitrogen). Real-time PCR was performed with SybrGreen Reagent (Qiagen, QuantiTect SYBR Green PCR Kit, Cat: 204143) and quantified by a CFX96 system (BioRAD, Inc.). For Fig. 3d, the specific loci L1M4 elements primers were designed and optimized based on ref. 27.

Embryoid body differentiation. For embryoid body differentiation experiment, feeder-free cultured ES cells were treated with 0.5% trypsin-EDTA free solution and resuspended with culture medium and counted. Then, cells were seeded at 200,000 cells per ml to Petri dishes with embryoid body differentiation medium (ESC medium without LIF and beta-ME). Medium was changed every 2 days.

Histone mass spectrometry. Histones were isolated in biological triplicate from wild-type and *Alkbh1* knockout cells by acid-extraction and resolved/visualized by SDS-PAGE/Coomassie staining. The low molecular weight region of the gel corresponding to core histones was excised and de-stained. The excised gel region containing the histones was treated with *d6*-acetic anhydride to convert unmodified lysine residues to heavy acetylated lysines (45 Da mass addition) as reported in ref. 41. Following *d6*-acetic anhydride treatment, the gel region was subjected to in-gel trypsin digestion. Histone peptides were analysed with a Thermo Velos Orbitrap mass spectrometer coupled to a Waters nanoACQUITY LC system as detailed in ref. 42. Tandem mass spectrometric data was searched with Mascot for the following possible modifications: heavy lysine acetylation, lysine acetylation, lysine monomethylation, lysine dimethylation and lysine trimethylation. For each biological replicate, histone H2A was identified with 100% sequence coverage across K118/119 that revealed predominately no detectable lysine methylation.

LC-MS/MS method for the determination of N⁶-methyladenine. DNA was digested with DNA Degradase Plus (Zymo Research) by following the manufacturer's instructions with small modification. Briefly, the digestion reaction was carried out at 37°C for 70 min in a 25 µl final volume containing 5 units of DNA Degradase Plus and 5 fMol of internal standard. Following digestion, reaction mixture was diluted to 110 µl and the digested DNA solution was filtered with a Pall NanoSep 3kDa filter (Port Washington, NY) at 8,000 r.p.m. for 15 min. After centrifugal filtration, the digested DNA solution was injected onto an Agilent 1200 HPLC fraction collection system equipped with a diode-array detector (Agilent Technologies, Santa Clara, CA). Analytes were separated by reversed-phase

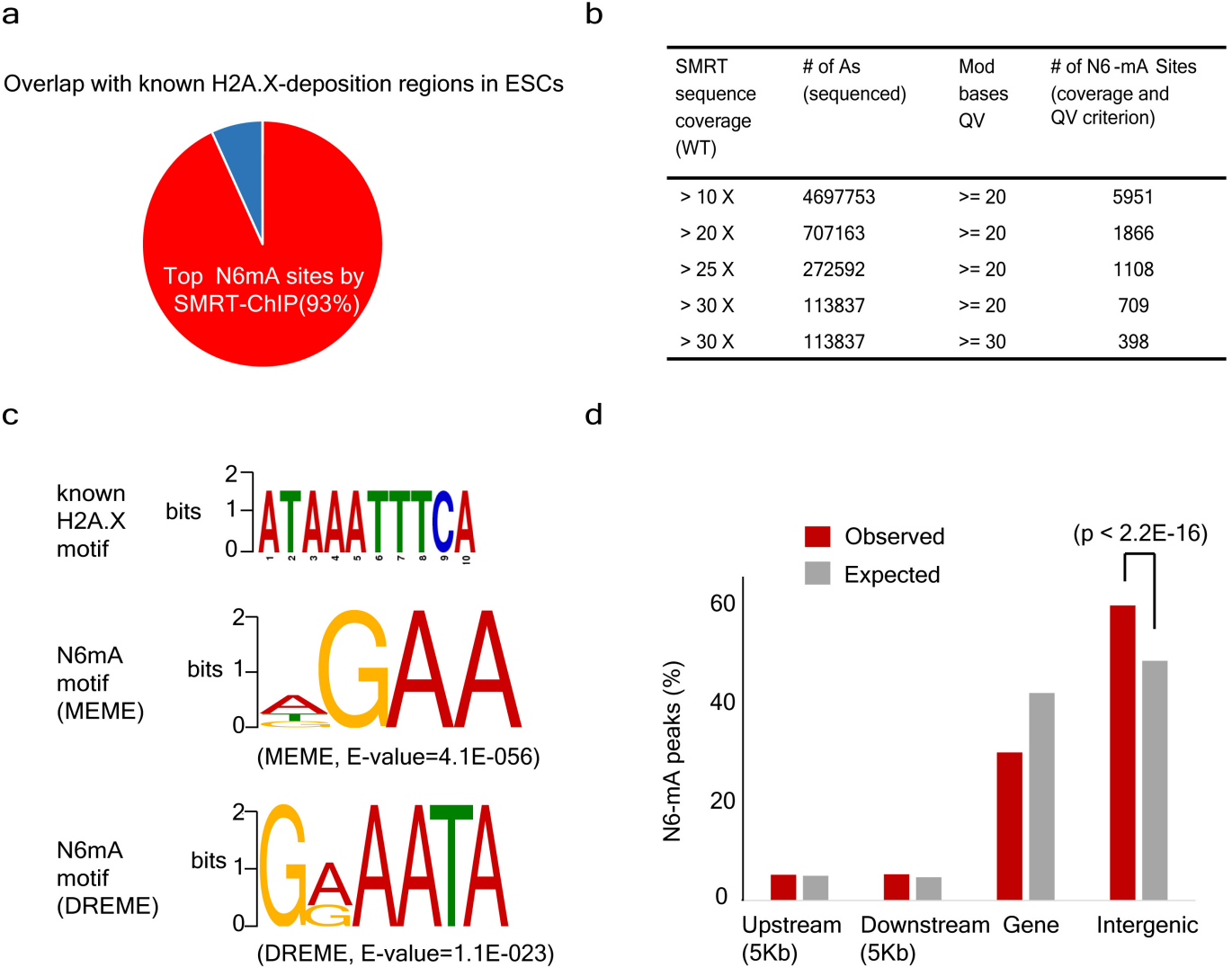
liquid chromatography using an Atlantis C₁₈ T3 (150 × 4.6 mm, 3 µm) column. The column temperature was kept at 30°C. For the purification of N⁶-mA, the mobile phases were water with 0.1% acetic acid (A) and acetonitrile with 0.1% acetic acid (B). The flow rate was 1.0 ml min⁻¹ with a starting condition of 2% B, which was held for 5 min, followed by a linear gradient of 4% B at 20 min, 10% B at 30 min, followed by 6 min at 80% B, then re-equilibration at the starting conditions for 20 min. dA and 6-Me-dA eluted with retention times of 14.7 and 27.0 min, respectively. The amount of dA in samples was quantitated by the UV peak area (λ = 254 nm) at the corresponding retention time using a calibration curve ranging from 0.2 to 5 nMol dA on column. For the simultaneous purification of N3-Me-dC, N1-Me-dA, N3-Me-dA, N6-Me-dA and dA, the mobile phases were water with 5 mM ammonium acetate (A) and acetonitrile (B). The flow rate was 0.45 ml min⁻¹ and the gradient elution program was set at following conditions: 0 min, 1% B; 2 min, 1% B; 40 min, 4% B; 60 min, 30% B; 65 min, 30% B; 65.5 min, 1% B, and 75 min, 1% B. N3-Me-dC, N1-Me-dA, N3-Me-dA, N6-Me-dA and dA eluted with retention times of 24.8, 25.0, 22.0, 60.2 and 54.2 min, respectively. The amount of dA in samples was quantitated by the UV peak area (λ = 254 nm) at the corresponding retention time using a calibration curve ranging from 0.9 to 7.2 nMol dA on the column. HPLC fractions containing target analyte were dried in a SpeedVac and reconstituted in 22 µl of D.I. water before LC-MS/MS analysis.

LC-MS-MS analysis of N3-Me-dC, N1-Me-dA, N3-Me-dA and N6-Me-dA was performed on Ultra Performance Liquid Chromatography system from Waters Corporation (Milford, MA) coupled to TSQ Quantum Ultra triple-stage quadrupole mass spectrometer (Thermo Scientific, San Jose, CA). 20 µl of sample was introduced into mass spectrometry through a 100 mm × 2.1 mm HSS T3 column (Waters) at flow rate of 0.15 ml/min. Mobile phases were comprised of water with 0.1% formic acid (A) or acetonitrile (B). Elution gradient condition was set as following: 0 min, 1% B; 3 min, 1% B; 15 min, 7.5% B; 15.5 min, 1% B; 20 min, 1% B. Ionization was operated in positive mode and analytes were detected in selected reaction monitoring (SRM) mode. Specifically, 6-Me-dA and its internal standard were detected by monitoring transition ions of *m/z* = 266.1 to *m/z* = 150.1 and *m/z* = 271.1 to *m/z* = 155.1, respectively. Similarly, N3-Me-dC, N1-Me-dA and N3-Me-dA was detected by monitoring transition ions of *m/z* = 242.1 to *m/z* = 126.1, *m/z* = 266.1 to *m/z* = 150.1 and *m/z* = 266.1 to *m/z* = 150.1, respectively. Mass spectrometry conditions were set as following: source voltage, 3,000 V; temperature of ion transfer tube, 280°C; skimmer offset, 0; scan speed, 75 ms; scan width, 0.7 *m/z*; Q1 and Q3 peak width, 0.7 *m/z*; collision energy, 17 eV; collision gas (argon), 1.5 arbitrary units. For quantification of N6-Me-dA, the linear calibration curves ranging from 1.5 to 750 fMol, were obtained using the ratio of integrated peak area of the analytical standard over that of the internal standard. The linear calibration curves for analysis of N3-Me-dC, N1-Me-dA and N3-Me-dA were obtained using integrated peak area of the analytical standard. N3-Me-dA is not commercial available and was prepared from the reaction between 3-methyladenine and deoxythymidine in the presence of nucleoside deoxyribosyltransferase II. The chemical identity of purified N3-Me-dA was confirmed by using an Agilent 1200 series Diode Array Detector (DAD) HPLC system coupled with Agilent quadrupole-time-of-flight (QTOF)-MS (Agilent Technologies, Santa Clara, CA). Electrospray ionization (ESI)-MS-MS spectrum of N3-Me-dA was obtained by in source fragmentation. One product ion was observed from MS/MS spectra of the protonated precursor ion of N³-Me-dA, resulting from the loss of the deoxyribosyl group. The accurate masses for parent and fragment ion are *m/z* = 266.1253 and *m/z* = 150.0774, with mass error 0.4 p.p.m. and 3.8 p.p.m., respectively. The method sensitivity for N3-Me-dC, N1-Me-dA, N3-Me-dA and N6-Me-dA was detected at 1.0 fmol, 1.6 fmol, 1.0 fmol and 1.6 fmol, respectively. In order to confirm the chemical identity of the N6-Me-dA isolated from HPLC purification, HPLC fractions containing N6-Me-dA was analysed by HPLC-QTOF-MS/MS. The chemical identity of N⁶-Me-dA in HPLC fractions was characterized on an Agilent 1200 series Diode Array Detector (DAD) HPLC system coupled with Agilent quadrupole-time-of-flight (QTOF)-MS (Agilent Technologies, Santa Clara, CA). HPLC separation was carried out on a C18 reverse phase column (Waters Atlantis T3, 3 µm, 150 mm × 2.1 mm) with a flow rate at 0.15 ml min⁻¹ and mobile phase A (0.05% acetic acid in water) and B (acetonitrile). The gradient elution program was set at following conditions: 0 min, 1% B; 2 min, 1% B; 15 min, 30% B; 15.5 min, 1% B; and 25 min, 1% B. N⁶-Me-dA was eluted with retention times of 12.7 min. The electrospray ion source in positive mode with the following conditions were used: gas temperature, 200°C; drying gas flow, 12 litres per min; nebulizer, 35 psi; Vcap, 4000 V; fragmentor, 175 V; skimmer, 67 V. Electrospray ionization (ESI)-MS-MS spectrum of N6-Me-dA isolated from genomic DNA was obtained by in source fragmentation. One product ion was observed from MS/MS spectra of the protonated precursor ion of N6-Me-dA, resulting from the loss of the deoxyribosyl group. The accurate masses for parent and fragment ion are *m/z* = 266.1245 and *m/z* = 150.0775, with mass error 3.0 p.p.m. and 3.1 p.p.m.,

respectively. The same MS/MS fragmentation spectra was obtained from analytical standard of N⁶-Me-dA.

For *in vitro* demethylation assay, sample was treated with EDTA to remove Fe²⁺. The mixture was transferred to Amicon Ultra Centrifugal Filter (EMD Millipore Corporation, 10K MWCO), followed by spin at 11,000 r.p.m. and 4 °C for 14 min. The concentrated sample was wash three times by adding 500 µl DI-H₂O, followed spin at 11,000 r.p.m. and 4 °C for 14 min. The washed sample was digested with DNA Degradase Plus (Zymo Research) by following manufacturer's instruction with small modification. Briefly, the digestion reaction was carried out at 37 °C for 60 min in 60 µl final volume containing 0.17 units per µl of DNA Degradase Plus and 50 fmol of Internal Standard of N⁶-Me-dA. Following digestion, reaction mixture was filtered with a Pall NanoSep 3kDa filter (Port Washington, NY) at 10000g and room temperature for 10 min to remove enzyme. The LC-MS/MS conditions for the quantification of dA and N⁶-Me-dA were set the same as those for quantification of N⁶-Me-dA in *in vivo* samples. The linear calibration curves for quantification of dA and N⁶-Me-dA was obtained using the ratio of integrated peak area of the analytical standard over that of the internal standard of N⁶-Me-dA.

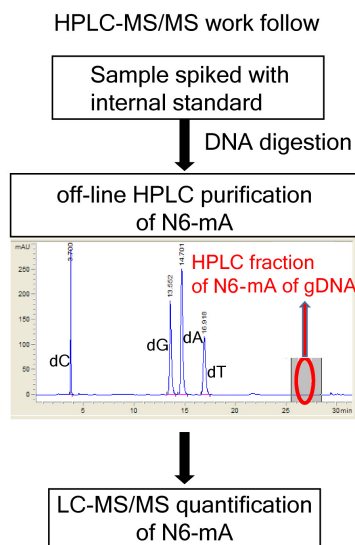
34. Tomomori-Sato, C., Sato, S., Conaway, R. C. & Conaway, J. W. Immunoaffinity purification of protein complexes from mammalian cells. *Methods Mol. Biol.* **977**, 273–287 (2013).
35. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**, 461–465 (2010).
36. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
37. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics* **25**, 1952–1958 (2009).
38. Song, Q. & Smith, A. D. Identifying dispersed epigenomic domains from ChIP-seq data. *Bioinformatics* **27**, 870–871 (2011).
39. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
40. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578 (2012).
41. Tackett, A. J. *et al.* I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J. Proteome Res.* **4**, 1752–1756 (2005).
42. Byrum, S. D., Taverna, S. D. & Tackett, A. J. Purification of a specific native genomic locus for proteomic analysis. *Nucleic Acids Res.* **41**, e195 (2013).



Extended Data Figure 1 | Low N6-mA levels in adult tissues and the lack of DNA alkylation adducts in ES cells. **a**, A majority of N6-mA peaks identified by SMRT-ChIP is located in H2A.X deposition region in ESCs determined by native ChIP. **b**, Number of SMRT-ChIP N6-mA sites at different coverage and QV cut-off. **c**, Top: A DNA motif of H2A.X

deposition region determine with standard ChIP-seq. Bottom: sequence motifs for N6-mA peaks at H2A.X deposition regions determined with SMRT-ChIP. **d**, Distribution of N6-mA peaks at H2A.X deposition regions (P value determined by binomial test).

a



b

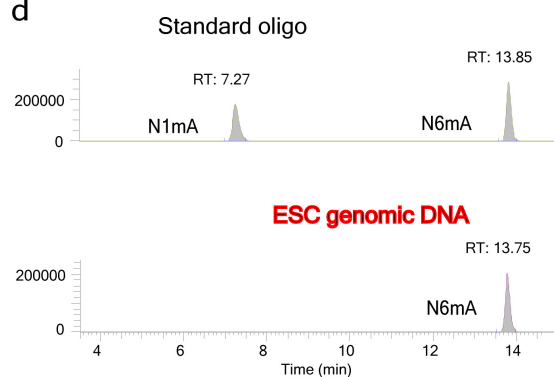
Tissue	Modification	Level (ppm of dA)
Mouse brain	N6-mA	2.4
Mouse kidney	N6-mA	1.9
Mouse thymus	N6-mA	3.8

c

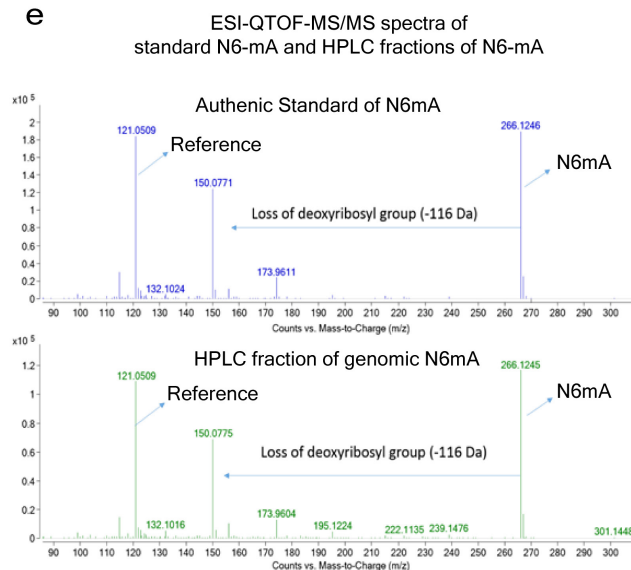
	N3mC	N1mA	N3mA
WT ESC	ND	ND*	ND
KO ESC	ND	ND	ND

(* ppm of dA; "ND", not detected: lower than quantification limit of LC-MS/MS)

d

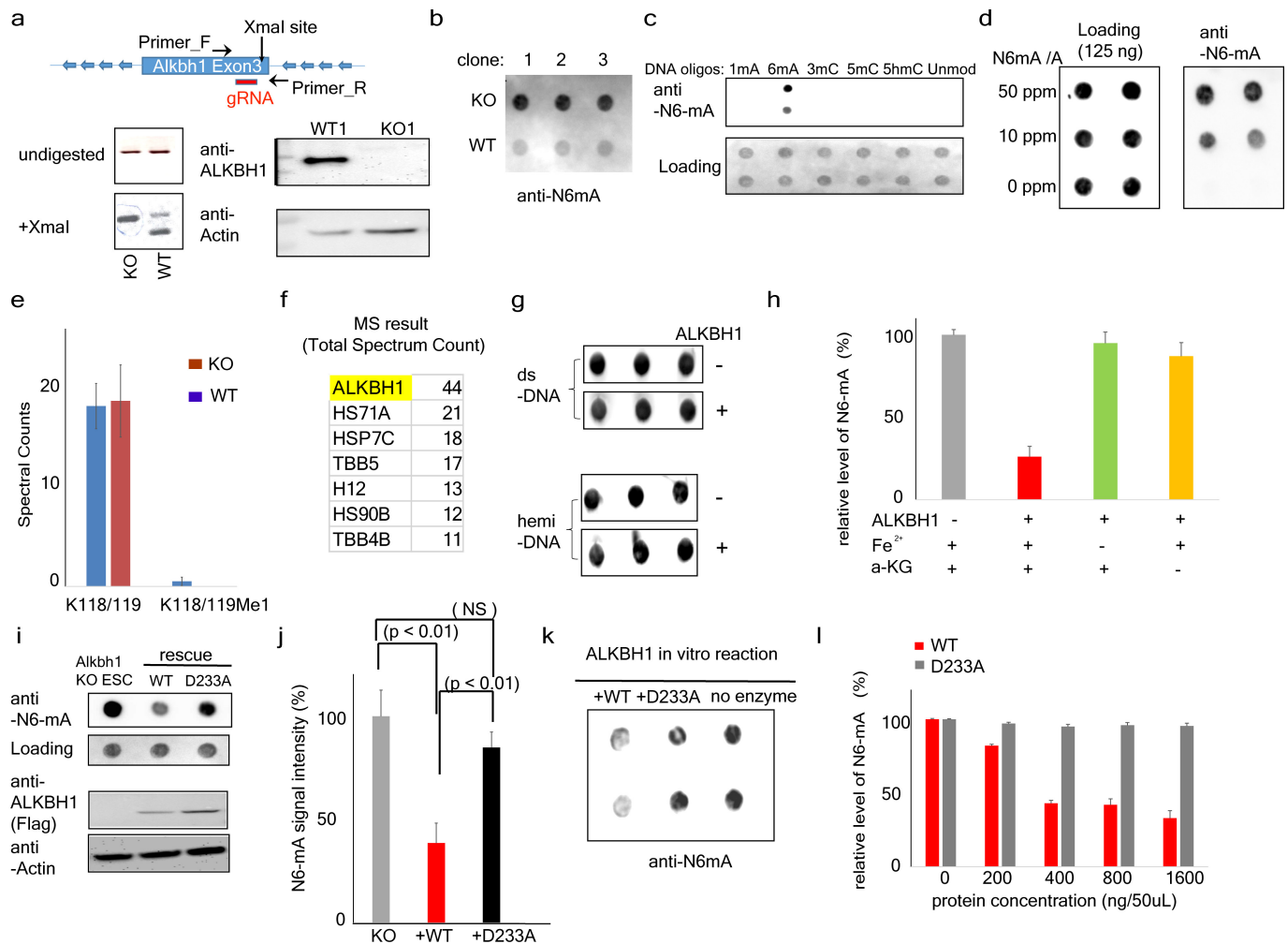


e



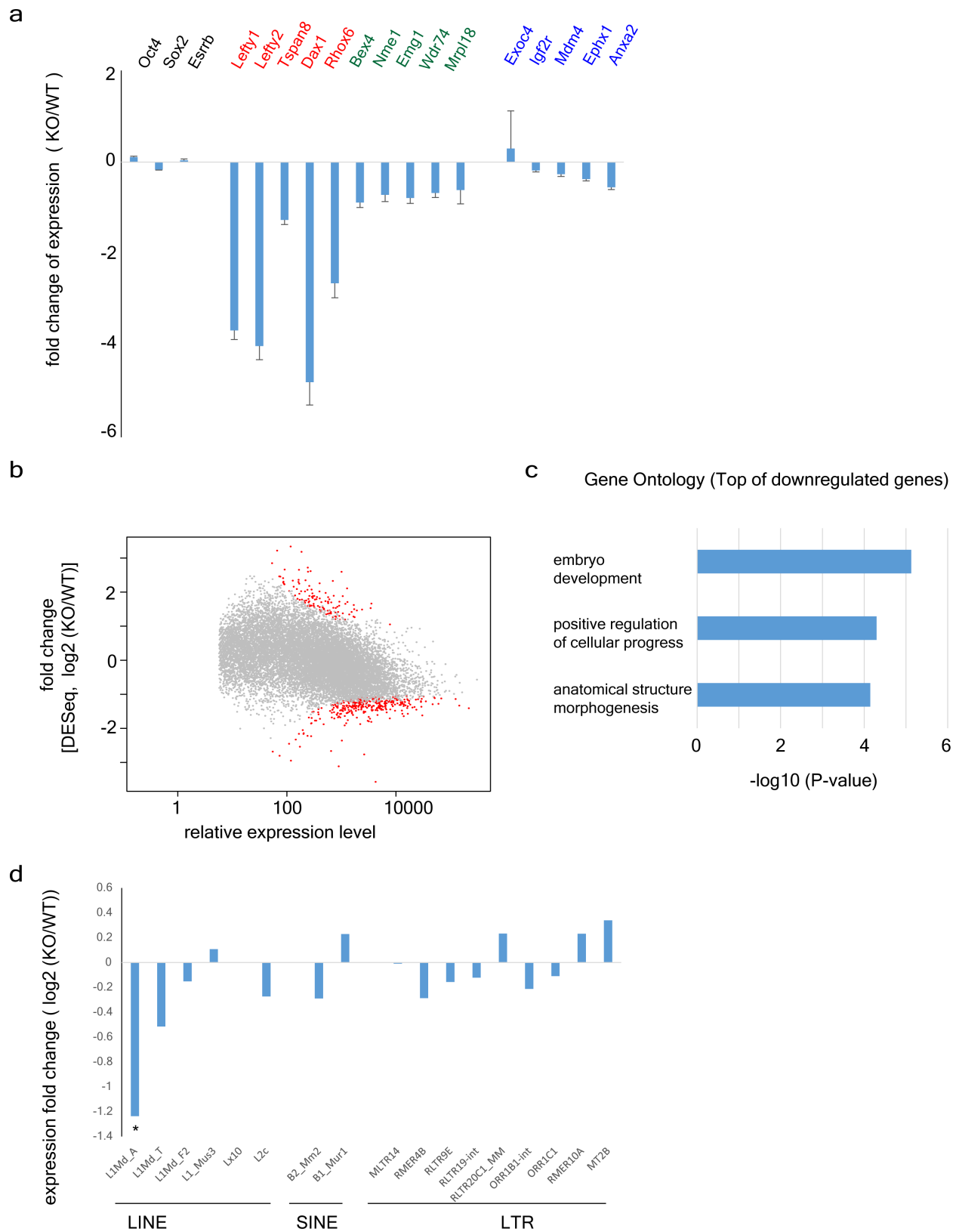
Extended Data Figure 2 | LC-MS/MS data of N6-mA. a, Experimental workflow for determining N6-mA level with LC-MS/MS. [^{15}N 5]N6-mA was used as the internal standard. b, N6-mA levels are ultralow in adult tissues. c, No detection of DNA alkylation adducts, such as N1-mA, N3-mA or N3-mC in mouse ES cells or *Alkbh1* knockout cells by MS.

d, LC-MS/MS analysis of N1-mA or N6-mA digested from synthetic oligonucleotides (top) and ES cell DNA samples (bottom). e, ESI-QTOF-MS/MS spectra of analytical standard of N6-mA nucleosides (top) and N6-mA containing HPLC fraction from ES cells.



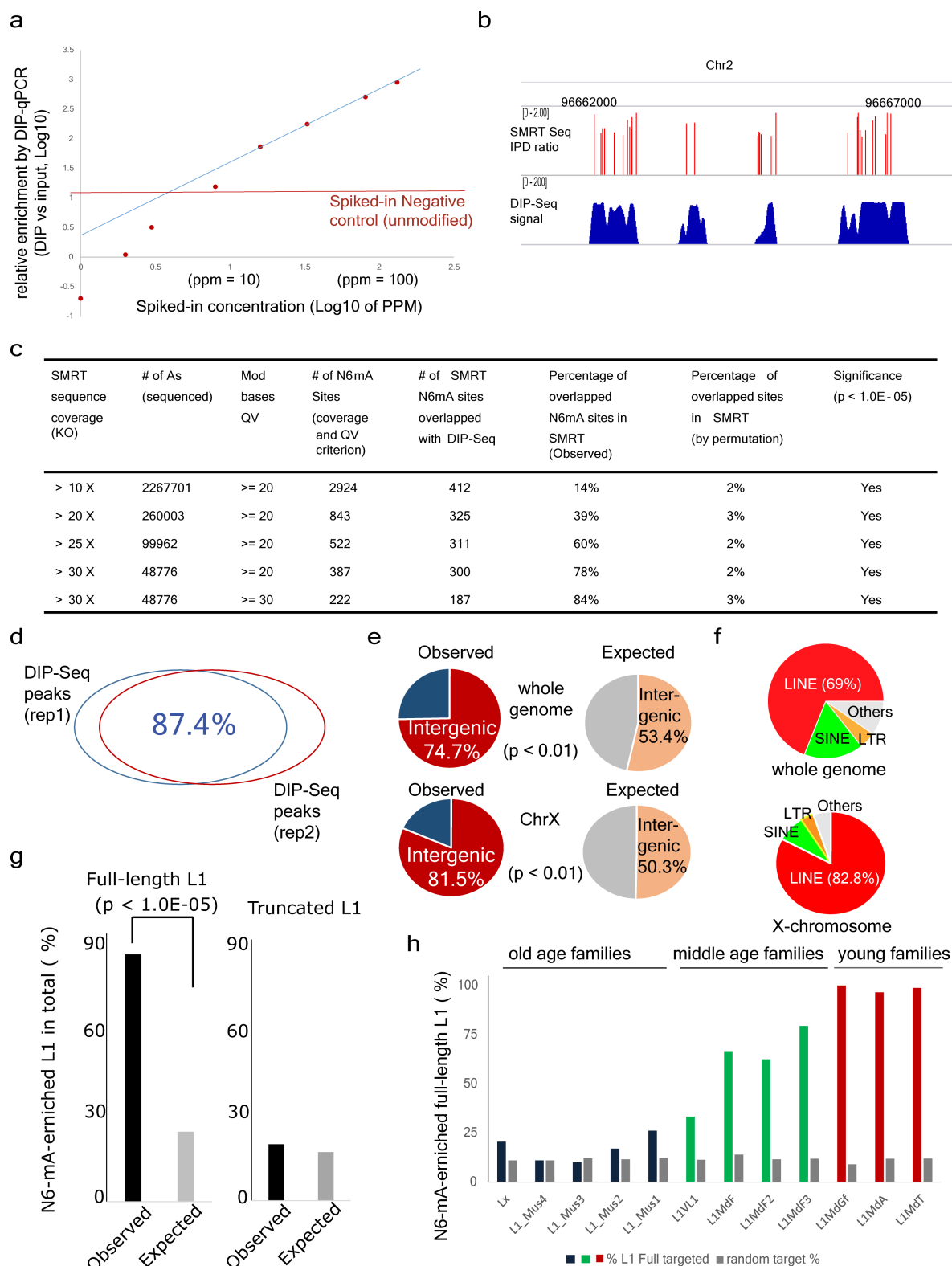
Extended Data Figure 3 | *Alkbh1* is a specific N6-mA demethylase *in vivo* and *in vitro*. **a**, Top: schematic of the CRISPR-Cas9 approach. *Alkbh1* KO alleles don't contain the XmaI site at exon 3. Bottom left: PCR-DNA digestion approach indicating the homozygosity of the knockout alleles, which are resistant to XmaI digestion. Bottom right: western blotting did not detect any ALKBH1 proteins in the KO cells. **b**, Three additional *Alkbh1* knockout ES cell clones show similar levels of N6-mA upregulation. Shown are dot blot results. **c**, Validating the specificity of anti-N6-mA antibodies with synthetic oligonucleotides. **d**, Validating the specificity of anti-N6-mA antibodies with DNA samples of different N6-mA/dA ratio. 125 ng of genomic DNA (MEFs) which does not contain any endogenous N6-mA was spiked with N6-mA containing oligonucleotides at the indicated concentration. **e**, Tandem mass spectrometric analysis shows the lack of H2AK118/119 methylation in wild-type or *Alkbh1* knockout ES cells. Spectral counts for H2A peptides containing K118/119 revealed that H2AK118/119 is predominately non-methylated at similar levels between wild-type and *Alkbh1* knockout ES cells. Spectral counts are reported as an average

with standard deviation from biological triplicate analyses. K118/119: no methylation; K118/119me1: K118/119 monomethylation. **f**, MS analysis showed that the co-purified factors with recombinant ALKBH1 proteins are mainly heat shock proteins. **g**, ALKBH1 proteins don't have noticeable activities towards to dual- or hemi-methylated double-stranded oligonucleotide substrates. **h**, ALKBH1 activities are dependent on Fe^{2+} and $\alpha\text{-KG}$. Error bars: standard deviation of triplicates. **i**, Ectopic expression of wild-type, but not mutant, *Alkbh1* (D233A) at the catalytic motif, can rescue the aberrant increase of N6-mA level in *Alkbh1* knockout ES cells. The wild-type and mutant *Alkbh1* were expressed at similar levels. **j**, Quantification of three independent rescue experiments in **i**. P value as labelled, determined by *t*-test; error bars, s.d. for three biological replicates. **k**, The demethylation activity of N6-mA by recombinant D233A mutant protein is much reduced in comparison with the wild-type counterpart. **l**, No significant activities were detected with increasing concentrations of recombinant D233A mutant proteins in demethylation reaction. Error bars, s.d. of triplicates.



Extended Data Figure 4 | RNA-seq analysis in *Alkbh1* knockout ES cells. **a**, RT-qPCR validation of the RNA-seq analysis. Unchanged genes (gene names labelled in black) identified by RNA-seq were unaltered in RT-qPCR analysis. Highly repressed (red), or modestly repressed (green) genes identified by RNA-seq also showed expected levels of repression in RT-qPCR analyses. Of note, the genes (blue) identified as upregulated in RNA-seq; however, they don't show differential expression (no significance) in RT-qPCR analysis, which further confirmed the

suppression function of ALKBH1. Error bars, s.d. of triplicates. **b**, MA plot of RNA-seq analysed by DESeq2, which shows the similar pattern to that of CuffDiff2 (see Fig. 3a and Methods). **c**, Gene ontology analysis demonstrated that lineage specifying factors involved in embryonic development are greatly downregulated by *Alkbh1* deficiency. **d**, RNA-seq transcripts of the representative subfamilies in three major retrotransposon superfamilies (LINE, SINE and LTR) in *Alkbh1* knockout ES cells (Methods).

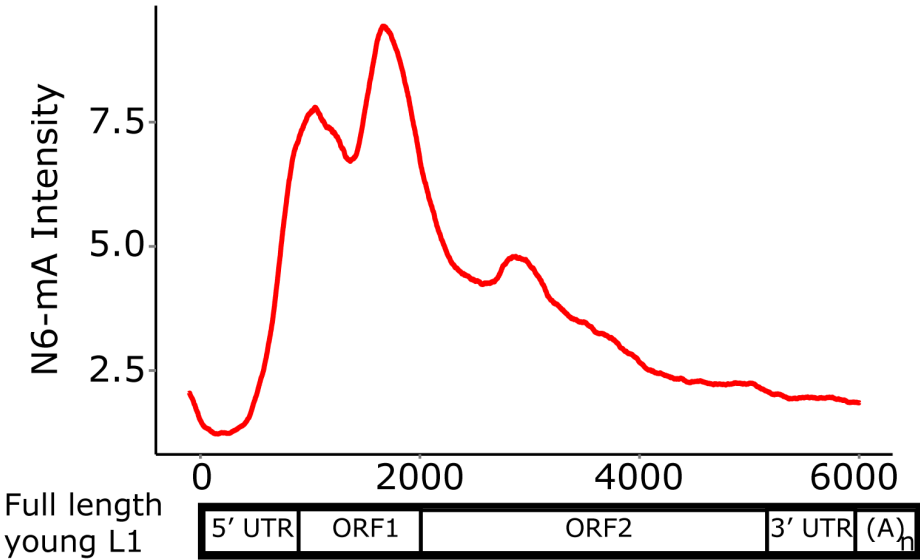


Extended Data Figure 5 | Validation of N6-mA DIP-seq approach.

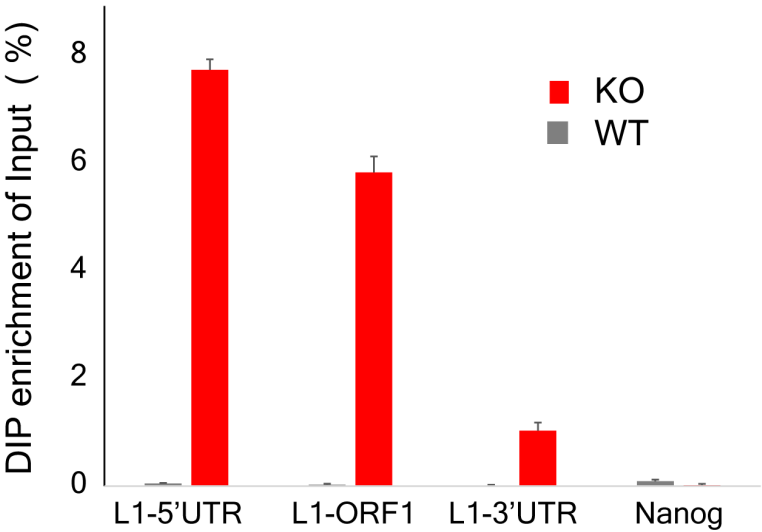
a, 'Spike-in' experiments for determining the threshold and linear response range of N6-mA DIP. Genomic DNAs were spiked with N6-mA containing oligonucleotides at indicated concentration (x axis). After N6-mA DIP, the relative enrichment of N6-mA over input control was determined by a RT-qPCR approach. Blue line: linear regression based on data points between 20–130 p.p.m. The threshold (the red line) is the background signals detected by RT-qPCR in which unmodified (control) oligonucleotides were spiked in. **b**, The track of different sequencing method showed N6-mA sites overlapped between SMRT-ChIP and DIP-Seq in *Alkbh1* knockout ES cells. **c**, Number of SMRT-ChIP N6-mA sites in *Alkbh1* knockout cells at different coverage

and QV cut-off. With rising coverage and QV cut-off, overlap between SMRT-ChIP N6-mA sites and DIP-Seq N6-mA sites also increases. **d**, The biological replicates of *Alkbh1* knockout ES cells N6-mA-DIP peaks show 87.4% overlap. **e**, A large majority of N6-mA peaks are in the intergenic regions at the whole-genome level or on the X chromosome. **f**, In *Alkbh1* knockout ES cells, N6-mA peaks are mainly targeted to LINE-1 transposons on the X chromosome or genome-wide. **g**, N6-mA peaks are significantly enriched on full-length, but not on truncated L1 elements ($P < 1.0 \times 10^{-5}$, chi-squared test). **h**, Enrichment of N6-mA in each full length L1 subfamily. Lx, L1_Mus1-4: >6 million years; L1V1, L1MdF1-4: 1.5–6 million years; L1MdGf, L1MdA, L1MdT: <1.5 million years.

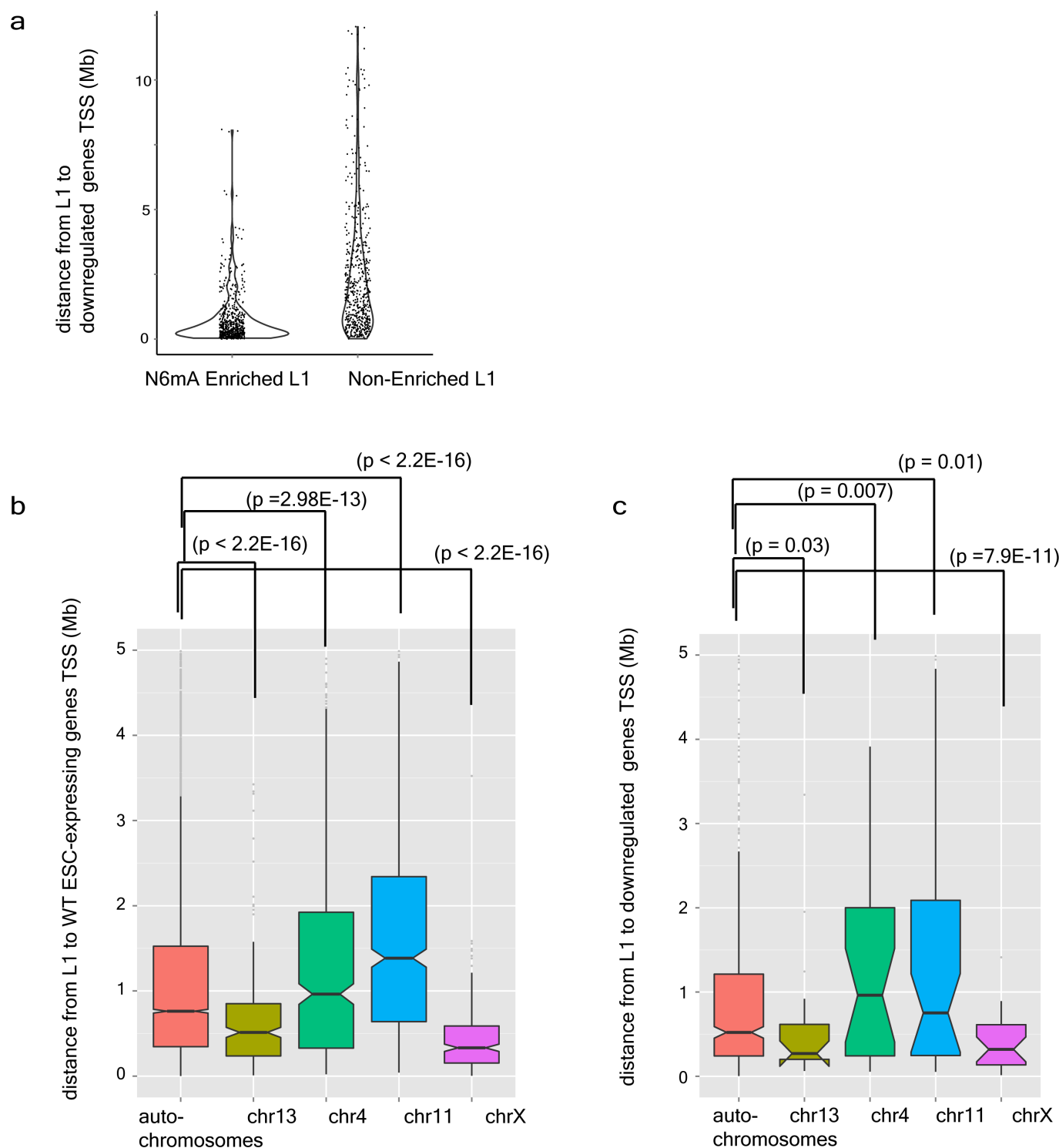
a



b

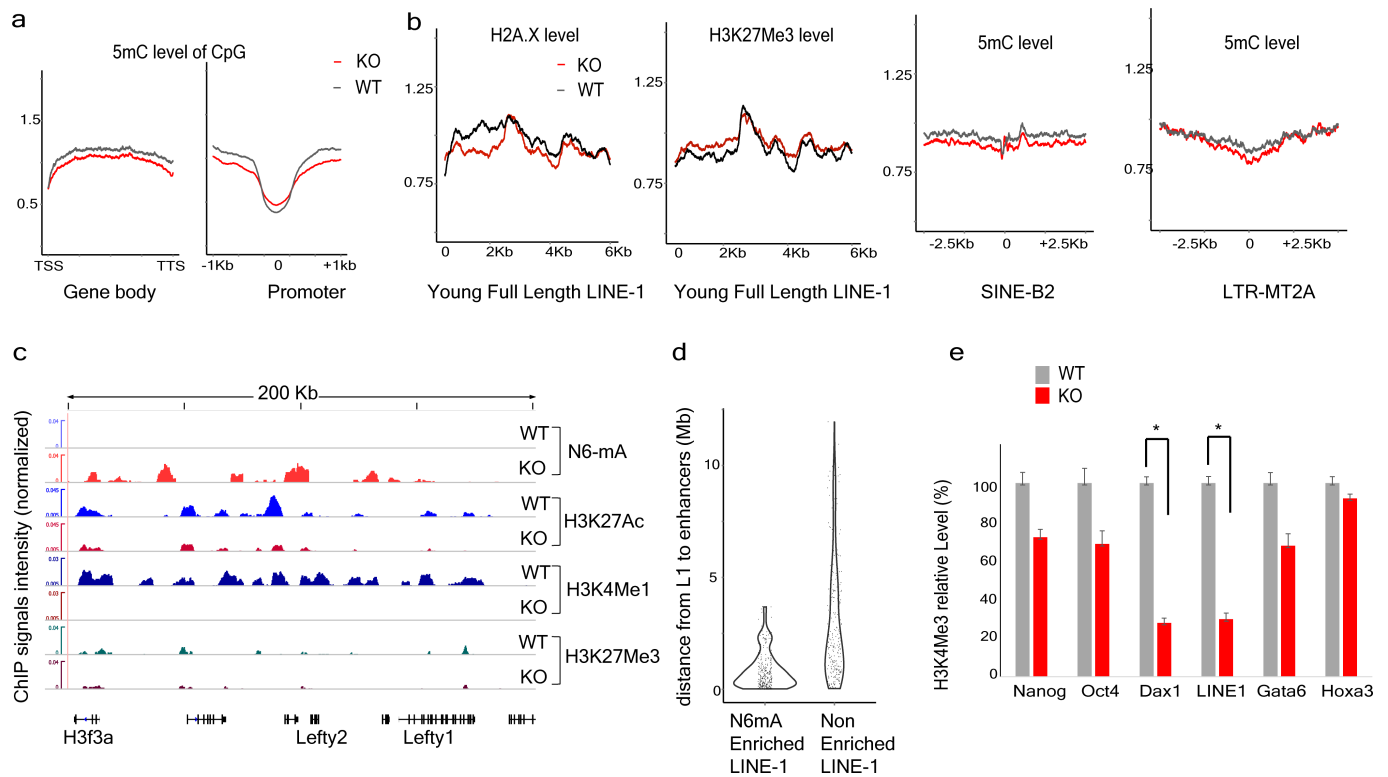


Extended Data Figure 6 | N6-mA enrichment on 5'-end of young full-length L1 elements. **a**, Aggregation plot shows that signal intensity of N6-mA at young full-length L1 is enriched at the 5' UTR and ORF1. **b**, qPCR analysis of N6-mA DIP samples confirmed the enrichment at the 5' UTR and ORF1 regions of L1 that are retained in the young full-length L1 elements, but not the 3' UTR or *Nanog* promoter.



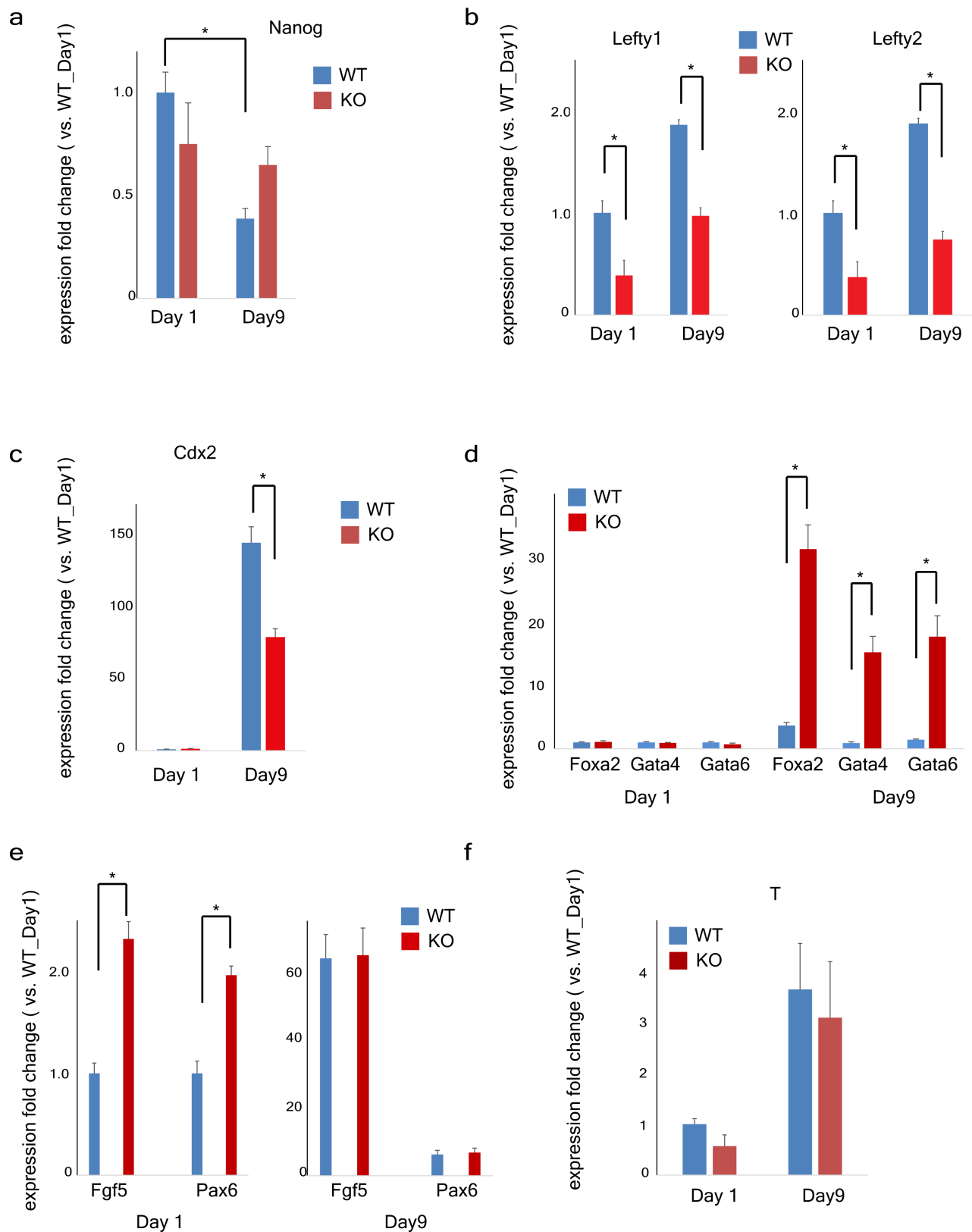
Extended Data Figure 7 | The correlation between N6-mA deposition on young full-length L1 elements and epigenetic silencing. **a**, Violin diagram of the density distribution of the distance between L1 and downregulated genes in *Alkbh1* knockout cells. **b**, The distances between

ES cells expressing genes in *Alkbh1* knockout ES cells and young full-length L1 elements were plotted for indicated chromosomes. **c**, The distances between downregulated genes in *Alkbh1* knockout ES cells and young full-length L1 elements were plotted for indicated chromosomes.



Extended Data Figure 8 | N6-mA accumulation correlates with epigenetic silencing. **a**, Normalized 5mC levels on gene bodies or promoters in wild-type or *Alkbh1* knockout ES cells. **b**, Histone marks (H2A.X or H3K27Me3) or 5 mC levels on young full-length L1 elements, SINE or LTR transposons. **c**, Representative sequencing tracks of decommissioned enhancers. H3K27Ac and H3K4me1 levels at this locus are greatly downregulated in *Alkbh1* knockout ES cells. See Supplementary

Table 2 for all decommissioned enhancers in *Alkbh1* knockout ES cells. **d**, Violin diagram shows the density distribution of the distance between L1 and decommissioned enhancers in *Alkbh1* knockout cells. **e**, ChIP-qPCR approach showed that H3K4me3 levels are decreased at the transcription start sites (TSS) of LINE-1 or Dax1, an X chromosome gene, while unchanged at the control gene TSS. * $P < 0.01$, *t*-test; error bars, \pm s.e.m. of three technical triplicates.



Extended Data Figure 9 | N6-mA accumulation results in imbalanced cell fate decisions during ESC differentiation. Wild-type or *Alkbh1* knockout ES cells were subject to embryoid body differentiation (Methods). mRNA samples were collected at day 1 or day 9. Gene expression levels were quantified by RT-qPCR approaches. * $P < 0.01$, t -test; error bars, \pm s.e.m. of technical triplicates. **a**, At day 9, *Nanog* expression is reduced significantly in wild-type ES-cell-derived embryoid bodies as expected, while its level in *Alkbh1* knockout ES-cell-derived embryoid bodies is still high. **b**, *Lefty-1* and *Lefty-2* are repressed at

day 1 or day 9 in *Alkbh1* knockout ES-cell-derived embryoid bodies. **c**, Activation of *Cdx2*, is insufficient in *Alkbh1* knockout ES-cell-derived embryoid bodies. **d**, However, expressions of other endoderm markers, *Foxa2*, *Gata4*, *Gata6*, are significantly higher in *Alkbh1* knockout ES-cell-derived embryoid bodies than wild-type ES-cell-derived embryoid bodies. **e**, Ectoderm markers, *Fgf5* and *Pax6* are transiently (day 1) overexpressed in *Alkbh1* knockout ES-cell-derived embryoid bodies. **f**, Mesoderm marker, *T/Brachyury* is similarly expressed in wild-type or *Alkbh1* knockout ES-cell-derived embryoid bodies during differentiation.

X-ray structures and mechanism of the human serotonin transporter

Jonathan A. Coleman¹, Evan M. Green^{1†} & Eric Gouaux^{1,2}

The serotonin transporter (SERT) terminates serotonergic signalling through the sodium- and chloride-dependent reuptake of neurotransmitter into presynaptic neurons. SERT is a target for antidepressant and psychostimulant drugs, which block reuptake and prolong neurotransmitter signalling. Here we report X-ray crystallographic structures of human SERT at 3.15 Å resolution bound to the antidepressants (S)-citalopram or paroxetine. Antidepressants lock SERT in an outward-open conformation by lodging in the central binding site, located between transmembrane helices 1, 3, 6, 8 and 10, directly blocking serotonin binding. We further identify the location of an allosteric site in the complex as residing at the periphery of the extracellular vestibule, interposed between extracellular loops 4 and 6 and transmembrane helices 1, 6, 10 and 11. Occupancy of the allosteric site sterically hinders ligand unbinding from the central site, providing an explanation for the action of (S)-citalopram as an allosteric ligand. These structures define the mechanism of antidepressant action in SERT, and provide blueprints for future drug design.

Serotonin (5-hydroxytryptamine or 5-HT) modulates the activity of the central nervous system as well as processes throughout the body ranging from cardiovascular function to digestion, body temperature, endocrinology and reproduction¹. Discovered in the late 1940s as a signalling molecule, serotonin increases vasoconstriction after blood clotting, that is, serum-tone². In the brain, the raphe nuclei synthesize serotonin from tryptophan, and distribute serotonin via long projections that reach nearly every major brain region. Serotonin is released into the synaptic cleft between neurons, where it diffuses to activate serotonin receptors, a group of G-protein-coupled receptors and ligand-gated ion channels that participate in both excitatory and inhibitory neurotransmission and modulate the release of many neurotransmitters and hormones. Thus, serotonergic signalling influences neurological processes including sleep, mood, cognition, pain, hunger and aggression behaviours. The discovery that serotonin reuptake into nerve terminals is inhibited by the tricyclic antidepressant imipramine in a manner similar to norepinephrine (also known as noradrenaline) reuptake provided an initial clue that transport occurs by a related reuptake system^{3–5}. Prozac was introduced as one of the first selective serotonin reuptake inhibitors (SSRIs) for the treatment of depression and, subsequently, the serotonin transporter gene (*SERT*, also known as *SLC6A4*) was cloned and proven to be the target of SSRIs^{6,7}.

SERT is a member of the neurotransmitter sodium symporter (NSS) family of transporters, which also includes the dopamine (DAT) and norepinephrine (NET) transporters. NSSs are responsible for the sodium- and chloride-dependent reuptake of neurotransmitters, thus terminating signalling of the biogenic amines^{8,9}. The unbinding of inhibitors can be further modulated by serotonin¹⁰ and antidepressants¹¹ acting at an allosteric site. Several neurological conditions are associated with NSS dysregulation, including depression, anxiety disorder, attention-deficit hyperactivity disorder, epilepsy and Parkinson's disease^{8,9,12}. Pharmacological modulation of NSS function through the use of therapeutic drugs such as tricyclic antidepressants and SSRIs has been used to treat many psychiatric disorders¹³. Illicit drugs such as cocaine and methamphetamines block neurotransmitter reuptake and are commonly abused psychostimulants, diminishing the well-being of users and constituting a tremendous socioeconomic burden.

Knowledge of NSS structure has been guided, in part, by experiments on the bacterial orthologue LeuT, as well as by studies of the *Drosophila* DAT (dDAT). This previous work has shown that the NSS family of transporters contain an inverted-topological repeat of transmembrane helices (TM) 1–5 and TM6–TM10, a 'central' or primary binding site for substrate and ions approximately halfway across the membrane-spanning region of the transporters^{14–18} and, in the outward-open conformation, a large extracellular vestibule. Recently, structures of the invertebrate dDAT have provided insight into NSS pharmacology^{19–21}. Nevertheless, these studies fall short of defining the structural determinants responsible for the markedly diverse pharmacological profiles of NSSs, the allosteric mechanism of human SERT, and important characteristics of human transporters. Here we present structures of the human serotonin transporter in complex with two of the most widely prescribed antidepressants: (S)-citalopram and paroxetine. Structures of SERT illuminate the molecular features of SSRI inhibition and allosteric regulation, as well as structural elements not reported in transporters previously studied.

Thermostable SERT–Fab complex

Wild-type human SERT²² is unstable in detergent micelles and refractory to crystallization. We thus screened a panel of SERT mutants for enhanced thermostability using a high-throughput ligand binding assay²³, and by fluorescence-detection size exclusion chromatography²⁴. Two thermostabilizing mutations, Ile291Ala and Thr439Ser, were introduced into SERT, yielding the ts2 construct, stable in short-chain detergents. Using transporter protein isolated from baculovirus-transduced mammalian cells²⁵, together with a recombinant anti-SERT Fab, we obtained small crystals of a ts2–Fab–paroxetine complex that diffracted X-rays to 4.5 Å resolution. To improve crystal order, we included a third thermostabilizing mutation (Tyr110Ala), yielding the ts3 construct, which further improved stability and produced crystals of the Fab complex with either (S)-citalopram or paroxetine that diffracted X-rays to 3.15 Å resolution (Extended Data Tables 1 and 2). Whereas the wild-type transporter exhibits serotonin transport with a Michaelis constant (K_m) of $1.9 \pm 0.3 \mu\text{M}$ (mean \pm s.e.m.) and a maximal velocity (V_{max}) of $23 \pm 1 \text{ pmol min}^{-1}$, similar to reported values⁸, ts2 has a K_m

¹Vollum Institute, Oregon Health & Science University, Portland, Oregon 97239, USA. ²Howard Hughes Medical Institute, Oregon Health & Science University, Portland, Oregon 97239, USA.

[†]Present address: Graduate Group in Biophysics, University of California, San Francisco, San Francisco, California 94158, USA.

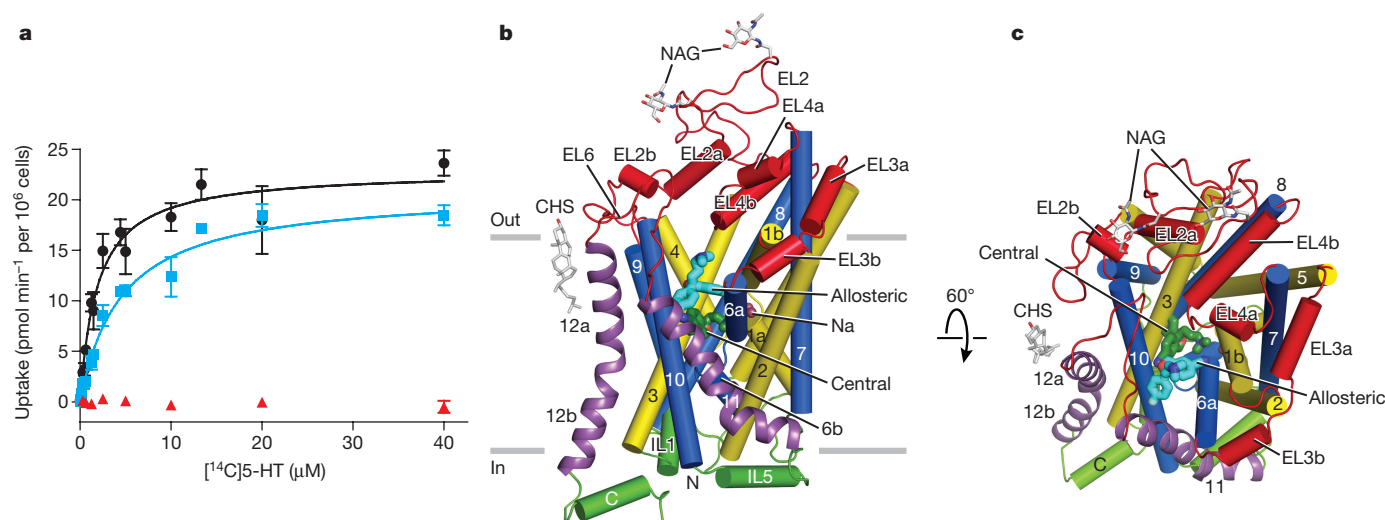


Figure 1 | Function and architecture of the human serotonin transporter. **a**, Michaelis–Menten plots of serotonin (5-HT) uptake by wild-type (black, circles), ts2 (blue, squares) and ts3 (red, triangles) transporters. Graph depicts an average of three independent experiments, each performed with triplicate measurements (error bars represent s.e.m.). **b**, Structure of SERT viewed parallel to the membrane. The (S)-citalopram

value of $4.5 \pm 0.6 \mu\text{M}$ and V_{max} value of $21 \pm 5 \text{ pmol min}^{-1}$ (Fig. 1a). No detectable transport activity was found for ts3.

Architecture of human SERT

The structure of human SERT bound to (S)-citalopram or paroxetine exhibits an outward-open conformation with the antidepressant drug bound to the central site, halfway across the membrane and wedged into a cavity made up of residues from TM1, TM3, TM6, TM8 and TM10 (Fig. 1b, c). A second (S)-citalopram molecule was found in the allosteric site, within the extracellular vestibule of the (S)-citalopram cocrystal structure, approximately 13 Å from the central site. Akin to dDAT and LeuT, SERT has 12 transmembrane-spanning helices (TM1–TM5 and TM6–TM10) related by a pseudo-two-fold axis^{14,16,20,21} (Extended Data Fig. 1). The ts2 and ts3 transporters superimpose well (Extended Data Table 3), demonstrating that the additional mutation of the ts3 construct does not substantially perturb the functionally active ts2 transporter structure (Extended Data Fig. 2a). TM1 and TM6 adopt short regions of non-helical conformation as they skirt the central ligand site and contribute residues that bind inhibitors as well as coordinate Na^+ and Cl^- ions. The conformations of TM1 and TM6 are incompatible with the formation of an occluded state, suggesting that the antidepressant molecules have locked the transporter in an outward-open conformation, similar to the inhibitor-bound outward-open conformations of dDAT and LeuT^{14,16,19,21,26} (Extended Data Table 3).

The extracellular surface of SERT is largely composed of extracellular loop (EL) 2, EL4 and EL6, with EL2 ‘combed-over’ the extracellular surface and providing 3,376 Å² of solvent-accessible surface area. A conserved disulfide bridge is formed between Cys200 and Cys209 in EL2 (ref. 27). EL2 is predicted to contain two N-linked glycosylation sites, Asn208 and Asn217 (ref. 28), and electron density for a N-acetylglucosamine moiety was found linked to Asn208; weak density was also found near Asn217. Similar to dDAT, the intracellular surface of the transporter is capped by intracellular loop (IL) 1, IL5 and the carboxy-terminal helix. Unlike LeuT, yet reminiscent of dDAT, TM12 has a pronounced kink halfway across the membrane. There is a cholesterol hemisuccinate (CHS) molecule bound near TM12a.

The crystal lattice packing between two SERT molecules occurs at the kink in TM12, which also overlaps with a two-fold axis of crystallographic symmetry (Extended Data Fig. 2c), thus generating an apparent SERT ‘dimer’. Experiments suggest that SERT is an oligomer

molecules at the central and allosteric site are shown as sticks in dark green and cyan, respectively. Sodium ions are shown as spheres in salmon. Cholesteryl hemisuccinate (CHS) and N-acetylglucosamine (NAG) are shown as sticks. **c**, View of SERT from the extracellular side of the membrane.

in the membrane²⁹. However, in detergent SERT is a monomer and we suggest that the SERT ‘dimer’ observed in this crystal form is unlikely to exist in a membrane bilayer because the predicted membrane-spanning regions of each protomer are not aligned with one another. Because the electron density for the Fab constant domain was poor, we also solved the structure of the Fab at 1.6 Å resolution to facilitate model building and refinement (Extended Data Fig. 2f and Extended Data Table 1). The Fab binds to a large extracellular surface consisting of EL2 and EL4 in a symmetry-related SERT, and this interface is further stabilized by interactions of EL2–EL2 and Fab–EL2 in the asymmetric unit (Extended Data Fig. 2d, e).

The structure of SERT shows that amino acid changes due to single nucleotide polymorphisms and mutations associated with psychiatric disorders are distributed throughout the structure (Extended Data Fig. 2b). Interestingly, most of the altered residues face solvent or lipid¹², thus rendering their effect on SERT structure and function obscure. Pro339Leu, however, is located in the non-helical region of TM6 neighbouring the ligand-binding site and, not surprisingly, this variant exhibits diminished transport activity. By contrast, other disease-associated mutations and polymorphisms, including mutations at Ile425 in TM8, Lys201Asn in EL2 (ref. 30) and Ser293Phe and Leu362Met in TM5 and TM7 enhance serotonin transport, respectively. Another class of mutations, Phe465Leu in TM9 and Leu550Val in TM11, probably destabilize the transporter or, as in the case of the Lys605Asn substitution in the C-terminal helix, render the transporter insensitive to protein kinase G regulation. With the establishment of SERT structural analysis, together with SERT expression and purification, we can now determine more precisely how these mutations alter the structure and activity of SERT.

Antidepressant bound at the central site

To probe the capacity of the ts2 and ts3 SERT constructs to bind antidepressants, we carried out binding studies using [³H](R/S)-citalopram and [³H]paroxetine (Fig. 2a, b). (R/S)-citalopram binds with dissociation constant (K_d) values of 2.1 ± 0.1 , 1.9 ± 0.3 and $2.9 \pm 0.5 \text{ nM}$ to the wild-type, ts2 and ts3 variants³¹, whereas paroxetine exhibits K_d values of 0.08 ± 0.03 , 0.17 ± 0.03 and $0.10 \pm 0.02 \text{ nM}$ for wild-type, ts2 and ts3 constructs³¹, respectively.

We next investigated the structure of the antidepressant binding site in the paroxetine and (S)-citalopram complexes by dividing the site into subsites A, B and C¹⁵ (Fig. 2c–e). Multi-crystal

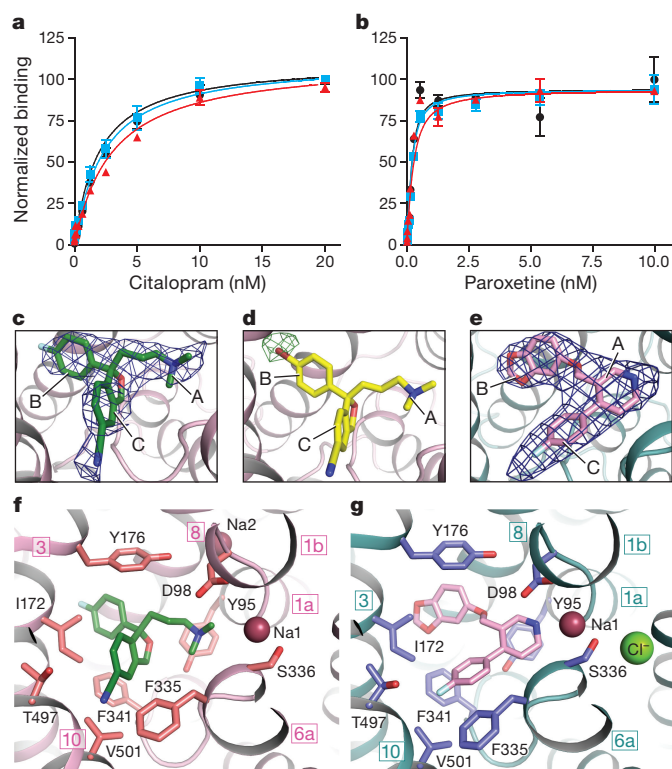


Figure 2 | Antidepressant binding and recognition. **a**, Graph of [^3H](*R/S*)-citalopram saturation binding to wild-type (black, circles), ts2 (blue, squares) and ts3 (red, triangles) transporters, showing the average of two independent experiments, each performed in triplicate (error bars represent s.e.m.). **b**, Plot of a [^3H]paroxetine saturation binding from a representative experiment (error bars represent s.e.m. from triplicate measurements). **c**, $F_o - F_c$ omit (*S*)-citalopram electron density (blue mesh), contoured at 3σ . The approximate positions of subsites A, B and C are shown. **d**, Anomalous difference electron density (green mesh), derived from Br-citalopram (yellow sticks) bound to the central site is shown (8.0σ contour level). **e**, $F_o - F_c$ omit electron density for paroxetine, contoured at 3σ . **f**, Interactions of (*S*)-citalopram (dark green) in the central binding site. **g**, Interactions of paroxetine (pink) with residues in the central binding site.

averaging of three (*S*)-citalopram data sets resulted in electron density maps that supported placement of the cyanophthalane group in subsite C and the fluorophenyl in subsite B, in agreement with detailed mutagenesis and ligand-binding studies³² (Fig. 2c). Nevertheless, because of the limited resolution of the diffraction data, we wanted to ensure that we had positioned (*S*)-citalopram correctly. To do this, we crystallized SERT with a bromine derivative of citalopram, (*R/S*)-1-[3-(dimethylamino)propyl]-1-(4-bromophenyl)-1,3-dihydroisobenzofuran-5-carbonitrile, in which the 4-fluoro group is replaced with a bromine atom (Br-citalopram). Upon analysis of the resulting anomalous difference Fourier map, we found a strong anomalous signal ($>9\sigma$) in subsite B, corresponding to the predicted position of the bromine atom of the Br-citalopram derivative and, by extension, the fluorine atom of (*S*)-citalopram (Fig. 2d). These data are consistent with the fluorophenyl group of (*S*)-citalopram occupying subsite B. The non-therapeutic *R*-enantiomer of citalopram has markedly weaker affinity for SERT, perhaps because the aromatic substituents swap subsites, relative to the *S*-enantiomer³³. Inspection of $F_o - F_c$ omit electron density maps allowed placement of paroxetine in the central binding site with the benzodioxol and fluorophenyl groups in subsites B and C (Fig. 2e). It is noteworthy that the chemically equivalent fluorophenyl groups of (*S*)-citalopram and paroxetine are positioned in different subsites.

The amine groups of (*S*)-citalopram and paroxetine occupy subsite A and interact with the carboxylate of the conserved Asp98 (ref. 34) at a distance of 4.1 and 3.1 Å (Fig. 2f, g), perhaps explaining, in part, why paroxetine has a higher affinity for SERT in comparison to (*S*)-citalopram. Tyr95 localizes 4.2 and 5.3 Å beneath the amine groups of (*S*)-citalopram and paroxetine, forming a cation- π interaction crucial for citalopram and mazindol potency³⁵. Tyr95 may also form a hydrogen bond with the oxygen of (*S*)-citalopram. Ser336 partners in an interaction network with ligands and ions by participating in Na^+ and Cl^- coordination, ions that are essential for ligand binding and substrate transport^{36–38}.

Subsite B is particularly important for high-affinity antidepressant interaction as evidenced by mutations that influence citalopram binding^{32,39}. Tyr176 engages in hydrophobic interactions with the fluorophenyl and benzodioxol groups of (*S*)-citalopram and paroxetine while also hydrogen bonding with Asp98. Ile172 and Phe341 define a non-polar ridge that cradles the hydrophobic groups of the drugs, and inhibitor binding is weakened upon mutation of these residues^{32,40}. Phe341 in SERT, which is equivalent to Phe325 in dDAT, has swung ‘downward’ by nearly 40° and forms an aromatic interaction with the ‘face’ of the cyanophthalane of (*S*)-citalopram and with the ‘edge’ of the fluorophenyl group of paroxetine (Extended Data Fig. 3a, b). Ser439, Leu443, Ala169 and Ala173 define a cavity that is more hydrophobic in SERT in comparison to the equivalent cavity in dDAT, and into which the fluorine and dioxol ring groups of (*S*)-citalopram and paroxetine are inserted (Extended Data Fig. 3c). The fluorophenyl group of (*S*)-citalopram is positioned 1.5 Å deeper into this space compared with the benzodioxol of paroxetine (Extended Data Fig. 3b).

The fluorophenyl group of paroxetine stacks parallel to the ring of Phe335 in subsite C. By contrast, for (*S*)-citalopram, the cyanophthalane forms an edge-to-face aromatic interaction. Phe335 defines the extracellular gate, and TM1 and TM6 are markedly different when comparing SERT to dDAT bound to a substrate analogue¹⁹ (Extended Data Table 3), showing that (*S*)-citalopram and paroxetine ‘prop’ TM6a in an outward-open conformation. Val501 and Thr497 form a mixed non-polar/polar surface into which the fluoro and cyano groups of paroxetine and (*S*)-citalopram are found. In the case of (*S*)-citalopram, the cyano group is inserted 2.1 Å further into subsite C, and the hydroxyl group of Thr497 is positioned 1.7 Å away from its position in the paroxetine state (Extended Data Fig. 3b). In accord with the SERT-citalopram X-ray structure, a modest increase in citalopram affinity is observed for the Thr497Ala mutant³⁹, which would allow additional space for the cyano group.

Ion-binding sites

Na^+ and Cl^- ions, which are essential for substrate transport and SSRI binding⁴¹, could be identified with electron densities $>3\sigma$ in $F_o - F_c$ ‘omit’ maps (Extended Data Fig. 4), at positions similar to those found in dDAT (Extended Data Table 4). The Na1 site is made up of residues contributed from TM1, TM6 and TM7 and the ion is coordinated by Ala96, Asn101, Ser336 and Asn368. Whereas Na1 is coordinated, in part, by a water molecule in dDAT, which in turn is hydrogen-bonded to Asp46, in SERT there is not sufficient density to place water at a similar position. The chloride ion is coordinated by Tyr121, Gln332, Ser336 and Ser372 from TM2, TM6 and TM7 with a mean coordination distance of 3.1 Å. Strong density for Na2 could be seen in the (*S*)-citalopram structure with the ion coordinated by Gly94, Val97, Leu434, Asp437 and Ser438 from TM1 and TM8. Placing ions in the omit densities led to a loss in $F_o - F_c$ density and the *B*-values of the ions match the values of surrounding residues. The mean coordination distance (2.4 Å) corresponds to known coordinate distances for sodium⁴². Only weak density for Na2 could be seen in the paroxetine structure, while in the (*S*)-citalopram complex the density for Cl^- was weak, perhaps reflecting the overall weaker density in these regions rather than a difference in occupancy.

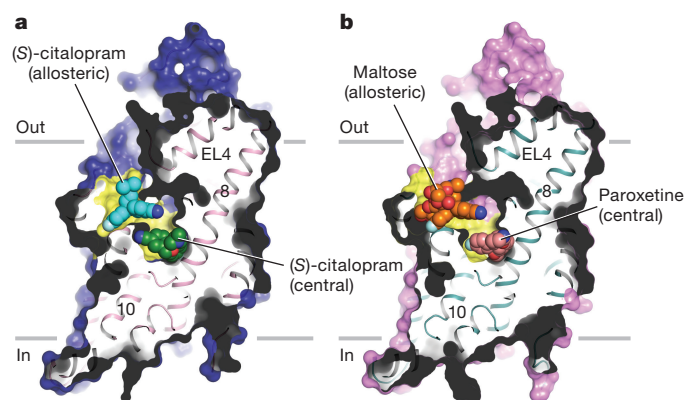


Figure 3 | Allosteric site. **a**, Sagittal slice through a surface representation of the (S)-citalopram-bound transporter. (S)-citalopram molecules bound to the allosteric (cyan) and central (green) sites are shown as spheres. **b**, A maltose head group (orange), derived from a detergent molecule and bound to the allosteric site, and paroxetine (pink), bound to the central site, are shown as spheres.

Extracellular and intracellular gates

The SERT–SSRI complexes adopt an outward-open conformation that exposes the cone-shaped extracellular vestibule to aqueous solution, providing a pathway for substrates, inhibitors and ions to reach the central binding site, approximately halfway across the membrane bilayer. As in LeuT, the extracellular vestibule contains residues that form the extracellular gate, and is lined by TM1b and TM6a, as well as by extracellular regions of TM3, TM8, TM10 and TM11, together with EL6 and the ‘tip’ of EL4. The mixed polar and non-polar character of the extracellular vestibule provides low affinity binding sites for small molecules, similar to LeuT, and in SERT we find electron density attributed to a second (S)-citalopram molecule in the (S)-citalopram cocrystal structure and a maltose detergent head group in the paroxetine complex within the extracellular vestibule^{14,43–45} (Fig. 3a, b).

At the base of the vestibule is the extracellular gate, and near the cytoplasmic face of SERT is the intracellular gate (Extended Data Fig. 5a, b). In SERT, Tyr176 and Phe335 define the lower portion of the extracellular gate and are separated by a distance of 10 Å, thus providing open access to the extracellular vestibule. In comparison to the extracellular gate of dDAT, the equivalent region in SERT exhibits notable structural changes: Tyr176 and Asp98 are separated by 4.0 Å and TM10 is closer to TM1b, bringing Glu494 and Arg104 within 4.8 Å, and thus the central site can only be accessed through the extracellular vestibule. The intracellular gate of SERT is closed, similar to the outward facing conformations of dDAT and LeuT, thus precluding direct access from the central ligand binding site to the intracellular solution (Fig. 3 and Extended Data Fig. 5b).

Allosteric site

To determine whether the off-rate of inhibitor from the central site is modulated by a ligand binding to an allosteric site in the ts3 construct, we measured the dissociation of [³H](R/S)-citalopram from the central site in the presence of saturating concentrations of cold (S)-citalopram. As shown in previous studies, micromolar concentrations of (S)-citalopram, serotonin and other ligands slow dissociation from the central site^{10,11}. For ts3, 100 μM (S)-citalopram decreased the first-order rate of [³H](R/S)-citalopram dissociation by nearly tenfold compared to buffer alone (0.0032 ± 0.0007 versus $0.025 \pm 0.002 \text{ min}^{-1}$) (Fig. 4a), with the wild-type and ts2 transporters exhibiting similar effects (wild-type: 0.004 ± 0.001 versus $0.035 \pm 0.004 \text{ min}^{-1}$; ts2: 0.0028 ± 0.001 versus $0.08 \pm 0.03 \text{ min}^{-1}$), thus showing that allosteric modulation of ligand unbinding is intact in the ts2 and ts3 constructs.

The allosteric binding site of (S)-citalopram is defined by residues in TM1b, TM6a, TM10 and TM11, and in EL4 and EL6 (Fig. 4b

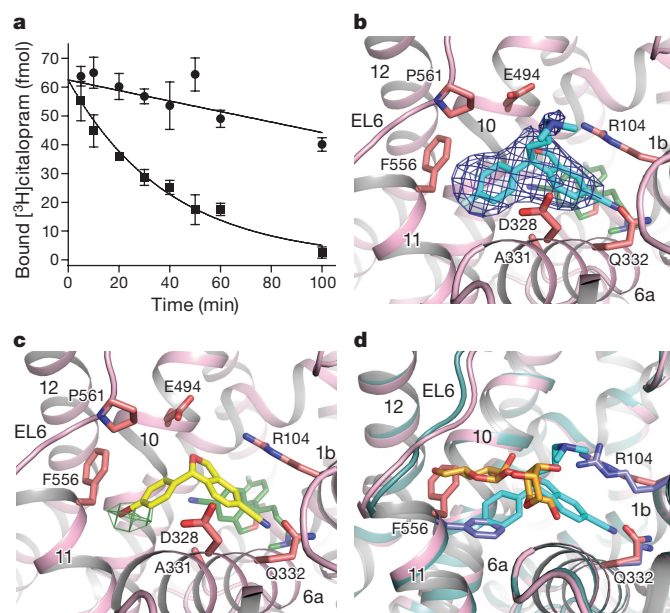


Figure 4 | Structural features of the allosteric site. **a**, Dissociation of [³H](R/S)-citalopram in the presence of buffer containing 100 μM (S)-citalopram (circles) or without ligand (squares). A representative experiment is shown (error bars represent s.e.m. from triplicate measurements). **b**, Allosteric site bound with (S)-citalopram (cyan). Residues in close proximity to (S)-citalopram are shown as sticks. A few atoms of (S)-citalopram at the central site (green sticks) are visible ‘below’ the (S)-citalopram molecule bound to the allosteric site. $F_o - F_c$ omit density of (S)-citalopram (blue mesh) in the allosteric site is shown (1.5σ contour level). **c**, Anomalous difference electron density (green mesh), derived from a Br-citalopram (yellow sticks) diffraction data set, is contoured at 5σ . **d**, Alignment of the allosteric site of the paroxetine (blue) and (S)-citalopram-bound (pink) structures. Maltose is in orange sticks. Superposition was performed over all C α atoms of the transporter.

and Extended Data Fig. 5c) with prominent electron density ($>5\sigma$ in $F_o - F_c$ omit maps) present in this region for crystals soaked with (S)-citalopram. Interestingly, mutagenesis of residues proximal to the allosteric site has been reported to severely alter allosteric potency⁴⁶ yet the physiological role of this site is not well established⁴⁷. Residues of the extracellular gate, Glu494 and Arg104, are located 4.1 and 4.8 Å from the aminopropyl group, while Asp328 is 6.8 Å away. Arg104 is also located 3.6 Å from the cyanophthalane ring and probably participates in a cation– π interaction, while the cyano group of the phthalane ring is 3.1 Å from the side-chain amide of Gln332. Ala331 forms a non-polar groove into which the ring system of (S)-citalopram is buttressed. Phe556 is 3.5 Å from the fluorophenyl group and participates in aromatic interactions while a proline repeat (Pro560–Pro561) in EL6 demarcates the upper portion of the allosteric site, 6.6 Å from the fluorophenyl entity.

To confirm the identity of the ligand bound to the allosteric site, we soaked crystals with Br-citalopram. A strong anomalous signal ($>5\sigma$) corresponding to bromine was detected in anomalous difference electron density maps, confirming the position and pose of citalopram in the allosteric site (Fig. 4c). However, there was not sufficient electron density to place the aminopropyl group of Br-citalopram and thus we excluded it from the structure.

In the paroxetine complex, we found electron density for a putative maltose entity, presumably derived from a detergent molecule (Extended Data Fig. 5d) occupying a position in the extracellular vestibule that partially overlaps with (S)-citalopram bound in the allosteric site (Fig. 4d). Upon analysis of the allosteric site of the (S)-citalopram and paroxetine complexes, we note considerable plasticity, presumably owing to the nature of the bound molecule. Relative

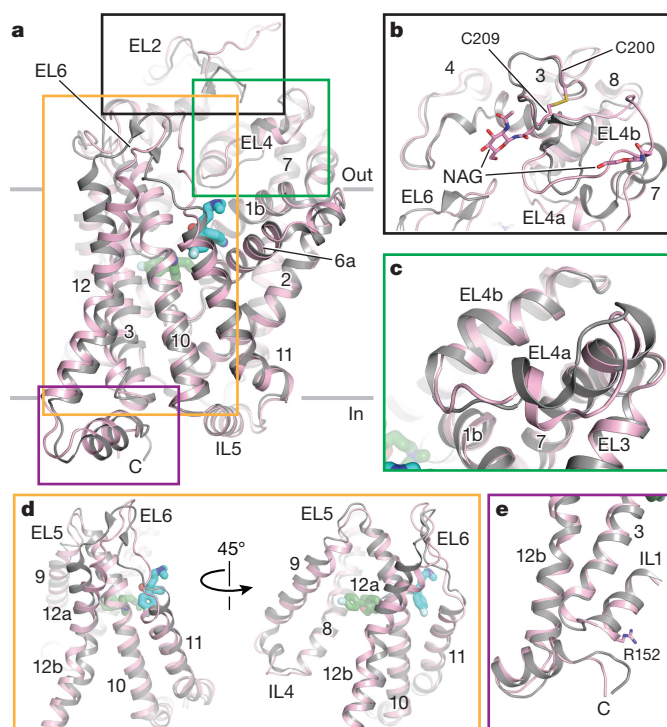


Figure 5 | Comparison of serotonin and dopamine transporters.

a, Overall alignment of SERT (pink) versus dDAT (grey) using TM1–TM12; regions in SERT with structural differences are boxed. (S)-citalopram bound to the central (green) and allosteric (cyan) sites shown as sticks. **b**, Close up view of EL2, *N*-acetylglucosamine (NAG; SERT) and the disulfide bridge between Cys200 and Cys209 are shown as sticks. **c**, View of EL4. **d**, Structural differences at the SERT allosteric site showing TM9, TM10, TM11, TM12, EL6 and IL4. **e**, Conformation of the C-terminal helix and IL1. Arg152 of SERT is shown as sticks.

to the (S)-citalopram-bound allosteric site, in the paroxetine structure Phe556 moves ‘downwards’ towards TM6a, to a position underneath the maltose. In addition, Arg104 moves 2 Å further into the allosteric site, while EL6 also moves 1.3 Å towards TM10, with the largest change occurring at Pro561. The malleability of the allosteric site opens the possibility that, depending on the shape and size of the allosteric ligand, occupancy of the allosteric site might not necessarily abrogate transport activity. Indeed, it is conceivable that there could be a spectrum of small molecules that range from inhibiting to enhancing transport activity.

Comparison of the allosteric site of SERT with the equivalent region of dDAT shows how the SERT site is distinct from that of dDAT, even though SERT and dDAT are highly similar in structure within their cores around the central ligand binding site (TM1–TM8; Extended Data Table 3). Indeed, there are marked differences between SERT and dDAT for TM9–TM12 and the extracellular loops (Fig. 5a and Extended Data Table 3). EL2, centrally positioned within the extracellular domain, is longer in SERT than in dDAT and participates in extensive interactions with EL4 and EL6, which together sculpt a portion of the allosteric site (Fig. 5b, c). Moreover, when comparing the amino acid sequences of human SERT, DAT and NET, EL6 displays one of the highest regions of diversity, consistent with the observation that allosteric ligands of SERT do not modulate DAT or NET. Furthermore, EL6 adopts a unique conformation not observed in dDAT because it extends more than 4 Å further towards EL2.

The conformation of TM9–TM12 also defines the allosteric site (Fig. 5d). Comparisons between dDAT and SERT illustrate that in SERT TM9 is shifted towards TM12, perhaps coordinated by contacts via EL5 and TM10, the latter of which contains a short stretch of π -helix near Glu494, a key residue of the extracellular gate. In SERT,

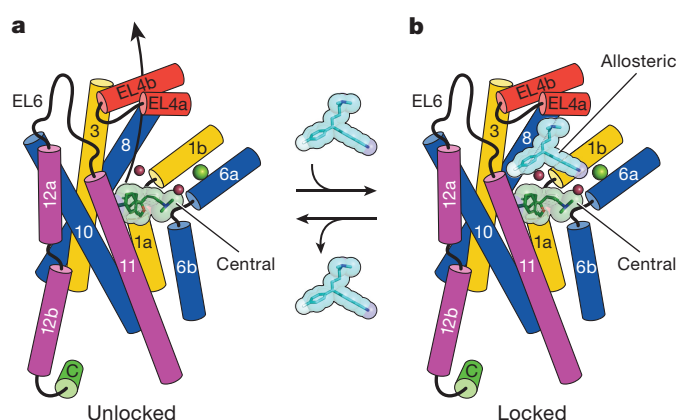


Figure 6 | Allosteric modulation of inhibitor binding. **a**, The SSRI (S)-citalopram (dark green) binds to the central site by wedging between scaffold helices 3, 8 and 10 and core helices 1 and 6. Sodium and chloride ions are shown as salmon and green spheres. **b**, (S)-citalopram (cyan) binds to the allosteric site made up of TM1b, TM6a, TM10, TM11, EL4 and EL6. Binding to the allosteric site slows dissociation of inhibitor from the central site.

TM11 extends further into the putative membrane environment in comparison to dDAT, thus providing a larger cavity for allosteric ligands, while TM12a splay inward to buttress TM10 and TM11. Finally, interaction of cholesterol, which is known to modulate transport and ligand binding⁴⁸ together with other lipid molecules, may reinforce the conformation of TM12. Indeed, in SERT a CHS molecule stacks against Trp573 in a groove formed by Leu577, Ile576 and Ala580 and the extracellular portion of TM12a (Extended Data Fig. 6b), along with a presumed alkyl chain of a detergent molecule bound in a cavity composed of residues from TM10 and TM12a (Extended Data Fig. 6c).

Intracellular surface and C-terminal hinge

IL5 and the intracellular half of TM11 are highly similar to dDAT, while IL4 is partially unwound due to the insertion of Trp458 (Fig. 5d). The C terminus of SERT mimics dDAT with a similar hinge and helix region (Fig. 5e). Glu615 is thought to form a salt bridge with Arg152 in IL1 (ref. 49), but no side-chain density is present, which makes assignment of C-terminal register not possible. We propose that the disorder of the C terminus is due to dynamic properties, perhaps related to its importance in trafficking⁵⁰.

Conclusion

The SERT–SSRI complexes capture the transporter in an inhibitor-bound, outward-open conformation, illustrating how the bulky ligands lodge in the central binding site, preventing substrate binding and transporter isomerization to occluded and inward-open conformations. Extensive interactions throughout the central binding site explain, in large part, the selectivity of SSRIs. The allosteric site is poised ‘above’ the central site, within the ‘walls’ of the extracellular vestibule, directly obstructing ligand egress from the central site, thus explaining how allosteric ligands slow the off-rate of inhibitors bound to the central site (Fig. 6). Taken together, the structures of the human serotonin transporter shed fresh insight into antidepressant recognition and the molecular basis for allosteric modulation of inhibitor binding and of transporter activity, thus providing a platform to design small molecules targeting the central and allosteric binding sites.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 November 2015; accepted 29 February 2016.

Published online 6 April 2016.

1. Berger, M., Gray, J. A. & Roth, B. L. The expanded biology of serotonin. *Annu. Rev. Med.* **60**, 355–366 (2009).
2. Rapport, M. M., Green, A. A. & Page, I. H. Serum vasoconstrictor, serotonin; isolation and characterization. *J. Biol. Chem.* **176**, 1243–1251 (1948).
3. Blackburn, K. J., French, P. C. & Merrills, R. J. 5-hydroxytryptamine uptake by rat brain *in vitro*. *Life Sci.* **6**, 1653–1663 (1967).
4. Carlsson, A., Fuxe, K. & Ungerstedt, U. The effect of imipramine on central 5-hydroxytryptamine neurons. *J. Pharm. Pharmacol.* **20**, 150–151 (1968).
5. Glowinski, J. & Axelrod, J. Inhibition of uptake of tritiated-noradrenaline in the intact rat brain by imipramine and structurally related compounds. *Nature* **204**, 1318–1319 (1964).
6. Hoffman, B. J., Mezey, E. & Brownstein, M. J. Cloning of a serotonin transporter affected by antidepressants. *Science* **254**, 579–580 (1991).
7. Blakely, R. D. *et al.* Cloning and expression of a functional serotonin transporter from rat brain. *Nature* **354**, 66–70 (1991).
8. Kristensen, A. S. *et al.* SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacol. Rev.* **63**, 585–640 (2011).
9. Brör, S. & Gether, U. The solute carrier 6 family of transporters. *Br. J. Pharmacol.* **167**, 256–278 (2012).
10. Wennogle, L. P. & Meyerson, L. R. Serotonin modulates the dissociation of [³H]imipramine from human platelet recognition sites. *Eur. J. Pharmacol.* **86**, 303–307 (1982).
11. Zhong, H. *et al.* An allosteric binding site at the human serotonin transporter mediates the inhibition of escitalopram by R-citalopram: kinetic binding studies with the ALI/VFL-SI/TT mutant. *Neurosci. Lett.* **462**, 207–212 (2009).
12. Hahn, M. K. & Blakely, R. D. The functional impact of SLC6 transporter genetic variation. *Annu. Rev. Pharmacol. Toxicol.* **47**, 401–441 (2007).
13. Andersen, J., Kristensen, A. S., Bang-Andersen, B. & Stromgaard, K. Recent advances in the understanding of the interaction of antidepressant drugs with serotonin and norepinephrine transporters. *Chem. Commun. (Camb.)* 3677–3692 (2009).
14. Singh, S. K., Piscitelli, C. L., Yamashita, A. & Gouaux, E. A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* **322**, 1655–1661 (2008).
15. Wang, H. *et al.* Structural basis for action by diverse antidepressants on biogenic amine transporters. *Nature* **503**, 141–145 (2013).
16. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
17. Singh, S. K. & Pal, A. Biophysical approaches to the study of LeuT, a prokaryotic homolog of neurotransmitter sodium symporters. *Methods Enzymol.* **557**, 167–198 (2015).
18. Kazmier, K. *et al.* Conformational dynamics of ligand-dependent alternating access in LeuT. *Nature Struct. Mol. Biol.* **21**, 472–479 (2014).
19. Wang, K. H., Penmatsa, A. & Gouaux, E. Neurotransmitter and psychostimulant recognition by the dopamine transporter. *Nature* **521**, 322–327 (2015).
20. Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structures of *Drosophila* dopamine transporter in complex with nixoxetine and reboxetine. *Nature Struct. Mol. Biol.* **22**, 506–508 (2015).
21. Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85–90 (2013).
22. Ramamoorthy, S. *et al.* Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. *Proc. Natl Acad. Sci. USA* **90**, 2542–2546 (1993).
23. Green, E. M., Coleman, J. A. & Gouaux, E. Thermostabilization of the human serotonin transporter in an antidepressant-bound conformation. *PLoS ONE* **10**, e0145688 (2015).
24. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
25. Goehring, A. *et al.* Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nature Protocols* **9**, 2574–2585 (2014).
26. Krishnamurthy, H. & Gouaux, E. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469–474 (2012).
27. Chen, J. G., Liu-Chen, S. & Rudnick, G. External cysteine residues in the serotonin transporter. *Biochemistry* **36**, 1479–1486 (1997).
28. Tate, C. G. & Blakely, R. D. The effect of N-linked glycosylation on activity of the Na⁺- and Cl⁻-dependent serotonin transporter expressed using recombinant baculovirus in insect cells. *J. Biol. Chem.* **269**, 26303–26310 (1994).
29. Kilic, F. & Rudnick, G. Oligomerization of serotonin transporter and its functional consequences. *Proc. Natl Acad. Sci. USA* **97**, 3106–3111 (2000).
30. Rasmussen, T. N., Plenge, P., Bay, T., Egebjerg, J. & Gether, U. A single nucleotide polymorphism in the human serotonin transporter introduces a new site for N-linked glycosylation. *Neuropharmacology* **57**, 287–294 (2009).
31. Owens, M. J., Knight, D. L. & Nemeroff, C. B. Second-generation SSRIs: human monoamine transporter binding profile of escitalopram and R-fluoxetine. *Biol. Psychiatry* **50**, 345–350 (2001).
32. Andersen, J. *et al.* Mutational mapping and modeling of the binding site for (S)-citalopram in the human serotonin transporter. *J. Biol. Chem.* **285**, 2051–2063 (2010).
33. Koldsø, H. *et al.* The two enantiomers of citalopram bind to the human serotonin transporter in reversed orientations. *J. Am. Chem. Soc.* **132**, 1311–1322 (2010).
34. Sørensen, L. *et al.* Interaction of antidepressants with the serotonin and norepinephrine transporters: mutational studies of the S1 substrate binding pocket. *J. Biol. Chem.* **287**, 43694–43707 (2012).
35. Barker, E. L. *et al.* High affinity recognition of serotonin transporter antagonists defined by species-scanning mutagenesis. An aromatic residue in transmembrane domain I dictates species-selective recognition of citalopram and mazindol. *J. Biol. Chem.* **273**, 19459–19468 (1998).
36. Forrest, L. R., Tavoulari, S., Zhang, Y. W., Rudnick, G. & Honig, B. Identification of a chloride ion binding site in Na⁺/Cl⁻-dependent transporters. *Proc. Natl Acad. Sci. USA* **104**, 12761–12766 (2007).
37. Kantcheva, A. K. *et al.* Chloride binding site of neurotransmitter sodium symporters. *Proc. Natl Acad. Sci. USA* **110**, 8489–8494 (2013).
38. Humphreys, C. J., Wall, S. C. & Rudnick, G. Ligand binding to the serotonin transporter: equilibria, kinetics, and ion dependence. *Biochemistry* **33**, 9118–9125 (1994).
39. Andersen, J. *et al.* Molecular determinants for selective recognition of antidepressants in the human serotonin and norepinephrine transporters. *Proc. Natl Acad. Sci. USA* **108**, 12137–12142 (2011).
40. Henry, L. K. *et al.* Tyr-95 and Ile-172 in transmembrane segments 1 and 3 of human serotonin transporters interact to establish high affinity recognition of antidepressants. *J. Biol. Chem.* **281**, 2012–2023 (2006).
41. Tavoulari, S., Forrest, L. R. & Rudnick, G. Fluoxetine (Prozac) binding to serotonin transporter is modulated by chloride and conformational changes. *J. Neurosci.* **29**, 9635–9643 (2009).
42. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L. & Minor, W. Data mining of metal ion environments present in protein structures. *J. Inorg. Biochem.* **102**, 1765–1776 (2008).
43. Singh, S. K., Yamashita, A. & Gouaux, E. Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448**, 952–956 (2007).
44. Zhou, Z. *et al.* LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science* **317**, 1390–1393 (2007).
45. Zhou, Z. *et al.* Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures. *Nature Struct. Mol. Biol.* **16**, 652–657 (2009).
46. Plenge, P. *et al.* Steric hindrance mutagenesis in the conserved extracellular vestibule impedes allosteric binding of antidepressants to the serotonin transporter. *J. Biol. Chem.* **287**, 39316–39326 (2012).
47. Jacobsen, J. P. *et al.* The interaction of escitalopram and R-citalopram at the human serotonin transporter investigated in the mouse. *Psychopharmacology (Berl.)* **231**, 4527–4540 (2014).
48. Scanlon, S. M., Williams, D. C. & Schloss, P. Membrane cholesterol modulates serotonin transporter activity. *Biochemistry* **40**, 10507–10513 (2001).
49. Koban, F. *et al.* A salt bridge linking the first intracellular loop with the C terminus facilitates the folding of the serotonin transporter. *J. Biol. Chem.* **290**, 13263–13278 (2015).
50. Susic, S. *et al.* Switching the clientele: a lysine residing in the C terminus of the serotonin transporter specifies its preference for the coat protein complex II component SEC24C. *J. Biol. Chem.* **288**, 5330–5341 (2013).

Acknowledgements We thank D. Cawley for generating monoclonal antibodies and Lundbeck for Br-citalopram. We thank A. Penmatsa and K. Wang for assistance with initial crystal screening, K. Dürr and W. Lü for help with Fab crystallization and structure refinement, respectively, L. Vaskalis for assistance with figures, H. Owen for help with manuscript preparation and other Gouaux laboratory members for discussions. We acknowledge the staff of the Berkeley Center for Structural Biology at the Advanced Light Source and the Northeastern Collaborative Access Team at the Advanced Photon Source for assistance with data collection. J.A.C. has support from a Banting postdoctoral fellowship from the Canadian Institutes of Health Research. We are particularly grateful to Bernie and Jennifer LaCroute for their generous support, as well as for funding from the National Institutes of Health (NIH) (5R37MH070039). E.G. is an Investigator with the Howard Hughes Medical Institute.

Author Contributions J.A.C., E.M.G. and E.G. designed the project. E.M.G. and J.A.C. developed thermostable constructs for crystallization. J.A.C. performed protein purification, crystallography and biochemical analysis. J.A.C., E.M.G. and E.G. wrote the manuscript.

Author Information The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under the following accession codes: ts3 paroxetine (5I6X), ts2 paroxetine (5I6Z), ts3 (S)-citalopram (5I71), ts3 (S)-citalopram (soaked) (5I73), ts3 Br-citalopram (5I74), ts3 Br-citalopram (soaked) (5I75), and 8B6 Fab (5I66). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (gouaux@ohsu.edu).

METHODS

No statistical methods were used to predetermine sample size.

Protein constructs. The cDNA encoding the human wild-type SERT²² was cloned into the BacMam vector²⁵ with a C-terminal GFP tag. For crystallization studies, the ts3 variant contained thermostabilizing mutations Y110A, I291A, T439S and mutation of surface-exposed cysteines C554A, C580A and C622A, and was fused to a C-terminal GFP followed by twin Strep [TrpSerHisProGlnPheGluLys(GlyGlyGlySer)₂GlyGlySerAlaTrpSerHisProGlnPheGluLys], and His10 purification tags. The His10 tag was used in thermostability studies to capture SERT on copper scintillation beads²³ while the twin Strep tag was used for large-scale purification. Thrombin cleavage sites (LeuValProArgGlySer) were introduced in the N and C terminus after residues Gln76 and Thr618. The ts2 construct is identical to ts3 except that ts2 does not have the Y110A mutation. To raise the 8B6 antibody, residues 73–616 of wild-type SERT were cloned into BacMam with a C-terminal StrepII tag (TrpSerHisProGlnPheGluLys) and without GFP.

Anti-human SERT antibody discovery and Fab expression. The 8B6 monoclonal antibody against SERT was raised by D. Cawley. StrepII-tagged SERT was purified by Strep Tactin affinity chromatography as described subsequently in DDM with 1 μ M paroxetine. Liposomes containing asolectin: cholesterol:lipid A:brain polar lipid (60:17:3:20) were prepared in TBS (20 mM Tris, pH 8, 100 mM NaCl) at a concentration of 40 mg ml⁻¹ by extrusion through 200-nm filters. Liposomes were saturated with 5 mM *n*-dodecyl- β -D-maltoside (DDM) and purified SERT was added to the detergent:lipid mixture. DDM was removed by three successive additions of 80 mg ml⁻¹ biobeads. For the first two additions, the biobeads were incubated for 2 h; the final incubation was overnight. Paroxetine (10 μ M) was added to the proteoliposomes after reconstitution. SERT-knockout mice were purchased from the Jackson Laboratory (mouse strain: 008355) and immunized with \sim 30 μ g of reconstituted SERT. Hybridoma cell lines were generated as described⁵¹ and screened by fluorescence-detection size-exclusion chromatography (FSEC)²⁴ and western blotting to select antibodies which recognize tertiary epitopes. The 8B6 monoclonal antibody was purified from hybridoma supernatants using 4-mercapto-ethyl-pyridine resin. Fab was purified from papain digested monoclonal antibody by cation exchange chromatography and was stored in 20 mM Tris, pH 8, 150 mM NaCl and 10% glycerol.

The sequences of the 8B6 Fab light and heavy chain genes were determined by standard techniques. The genes of the 8B6 Fab were cloned into a bicistronic insect cell expression vector, including a GP67 signal peptide. A thrombin cleavage site and 8His tag were fused to the C terminus of residues 1–235 of the heavy chain. The 8B6 Fab was purified from Sf9 supernatant by metal ion affinity chromatography followed by cation exchange chromatography.

Transporter expression and purification. The human SERT constructs were expressed as C-terminal GFP fusions using baculovirus-mediated transduction of mammalian HEK293S GnT1⁻ cells, as previously described^{25,52}. Cells were subsequently solubilized in 50 mM Tris, pH 8, 150 mM NaCl containing 20 mM DDM, 2.5 mM CHS, 0.5 mM dithiothreitol (DTT) in the presence of 1 μ M inhibitor (paroxetine, (S)-citalopram, or Br-citalopram). The lysate was passed over 10 ml of Strep Tactin resin, washed with 18 column volumes of 1 mM DDM, 0.2 mM CHS, 5% glycerol, 25 μ M lipid (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine, and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol at a molar ratio of 1:1:1), and 1 μ M ligand in TBS. SERT was eluted in the same buffer containing 5 mM desthiobiotin. The N and C termini containing GFP and purification tags were removed by thrombin digestion and *N*-linked sugars were truncated using EndoH. SERT was mixed with recombinant 8B6 Fab at a 1:1.2 molar ratio. In the case of Br-citalopram complexed at the central site, Fab purified from hybridoma cells was used. The resulting complexes were further purified by size-exclusion chromatography in TBS supplemented with 40 mM *n*-octyl β -D-maltoside, 0.5 mM CHS, 5% glycerol, 25 μ M lipid, and 1 μ M inhibitor. The purified SERT-8B6 complex was concentrated to 2 mg ml⁻¹ and the transporter solution was spiked with 10 μ M inhibitor and 1 μ M 8B6 Fab immediately before crystallization.

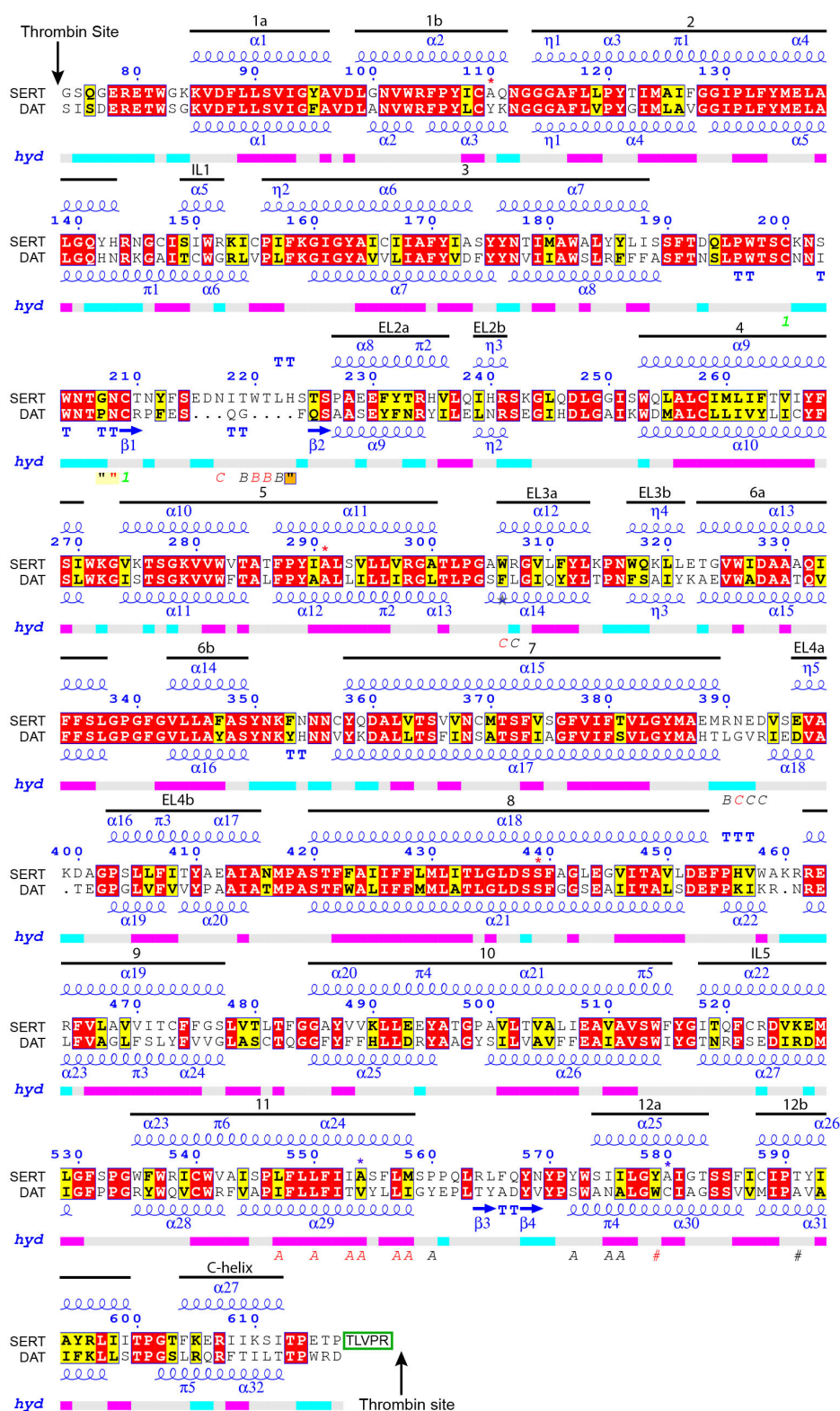
Crystallization. The SERT–Fab complex and Fab crystals were grown by hanging drop vapour diffusion. For the ts3–Fab complex, crystals appeared after several days under conditions with a reservoir solution composed of 100 mM Tris, pH 8.5, 25–125 mM KCl, 32.5–34% PEG 400, and 0.5% 6-aminohexanoic acid with either

paroxetine or (S)-citalopram at ratio of 2 μ l protein to 1 μ l of reservoir solution. Crystals of ts3–Fab complex with Br-citalopram bound to the central site were grown with a reservoir solution composed of 100 mM Tris, pH 8.5, 125 mM NaCl, 125 mM MgCl₂ and 33.4% PEG 400. ‘Soaked’ SERT–Fab crystals were prepared by addition of 5 mM ligand ((S)-citalopram or Br-citalopram) overnight to the crystals as indicated. For the ts2–Fab complex, crystals were grown using a reservoir solution containing 100 mM Tris, pH 8, 250 mM LiCl and 34.7% PEG 400 with paroxetine. All SERT–Fab crystals were grown at 4 °C, reaching full size after 14 days, and were directly flash-frozen in liquid nitrogen before X-ray diffraction data collection. The 8B6 Fab from hybridoma cells was crystallized in 100 mM HEPES, pH 7.5, 300 mM NH₄H₂PO₄ and 22.5% pure PEG (0.3–0.8 kDa PEG, Microlytic) at 4 °C. Fab crystals were cryoprotected with 25% ethylene glycol before flash cooling in liquid nitrogen.

Data collection and structure refinement. X-ray diffraction data sets were collected at the Advanced Photon Source (Argonne National Laboratory, beamline 24-IDC and 24-IDE) and at the Advanced Light Source (Lawrence Berkeley National Laboratory, beamline 5.0.2). X-ray data was processed using XDS⁵³. Molecular replacement was carried out to solve the 8B6 structure using the constant and variable domains of a preexisting Fab structure (PDB code 4LQF) as independent search models, followed by automated and manual model building and refinement using Phaser, Coot, Phenix AutoBuild and Phenix Refine^{54–56}. The SERT–Fab data were processed using the microdiffraction assembly method⁵⁷ where indicated. The SERT structure was solved by molecular replacement using a multi-model search in Phaser with a homology model built using the dDAT structure (PDB code 4M48) and Modeller⁵⁸ and the constant and variable domains of the 8B6 Fab structure. Several rounds of refinement and manual model building were carried out using Phenix and Coot, respectively, until the models were refined to acceptable R-factors and stereochemistry.

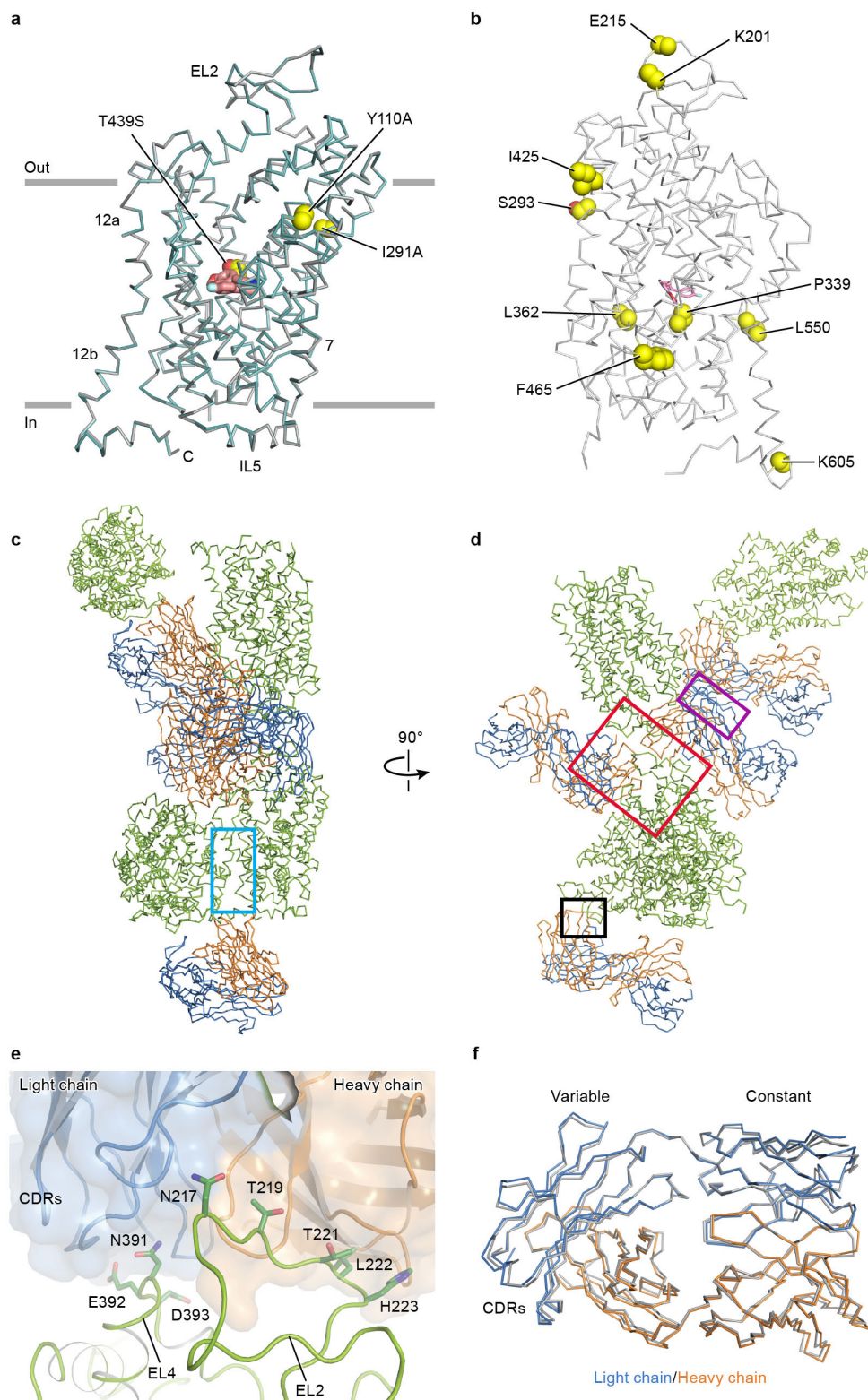
Radioligand binding and uptake assays. Ligand binding experiments were carried out by adding HEK293 membranes containing SERT to a final concentration of 2 nM in 1 ml of TBS with either [³H]paroxetine 0.01–10 nM or [³H](R/S)-citalopram 0.01–20 nM. Reactions were rotated at room temperature for 4 h followed by filtering through a glass microfibre filter prewet with 0.4% polyethylenimine in TBS. Membranes were washed three times with 4 ml of TBS followed by liquid scintillation counting. Data was fit to a single-site binding curve accounting for ligand depletion. For dissociation experiments, 20 nM SERT in membranes was mixed with 40 nM [³H](R/S)-citalopram in 10 μ l; samples were diluted to 1 ml in TBS with 100 μ M (S)-citalopram, or without ligand, followed by filtering. For uptake assays, \sim 1 \times 10⁵ HEK293 cells in 96-well Cytostar T plates were transfected with 0.2 μ g of plasmid with Polyjet. After 24–36 h, cells were washed with 25 mM HEPES-Tris, pH 7.0, 130 mM NaCl, 5.4 mM KCl, 1.2 mM CaCl₂, 1.2 mM MgSO₄, 1 mM ascorbic acid and 5 mM glucose. For a control, 10 μ M paroxetine was added. [¹⁴C]5-hydroxytryptamine at concentrations of 0.02–40.0 μ M was added and uptake was followed using a MicroBeta scintillation counter. Data were fit to a Michaelis–Menten equation.

- Galfre, G., Howe, S. C., Milstein, C., Butcher, G. W. & Howard, J. C. Antibodies to major histocompatibility antigens produced by hybrid cell lines. *Nature* **266**, 550–552 (1977).
- Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous *N*-glycosylation by a tetracycline-inducible *N*-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl Acad. Sci. USA* **99**, 13419–13424 (2002).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
- Bunkóczi, G. *et al.* Phaser.MRage: automated molecular replacement. *Acta Crystallogr. D* **69**, 2276–2286 (2013).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012).
- Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
- Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).



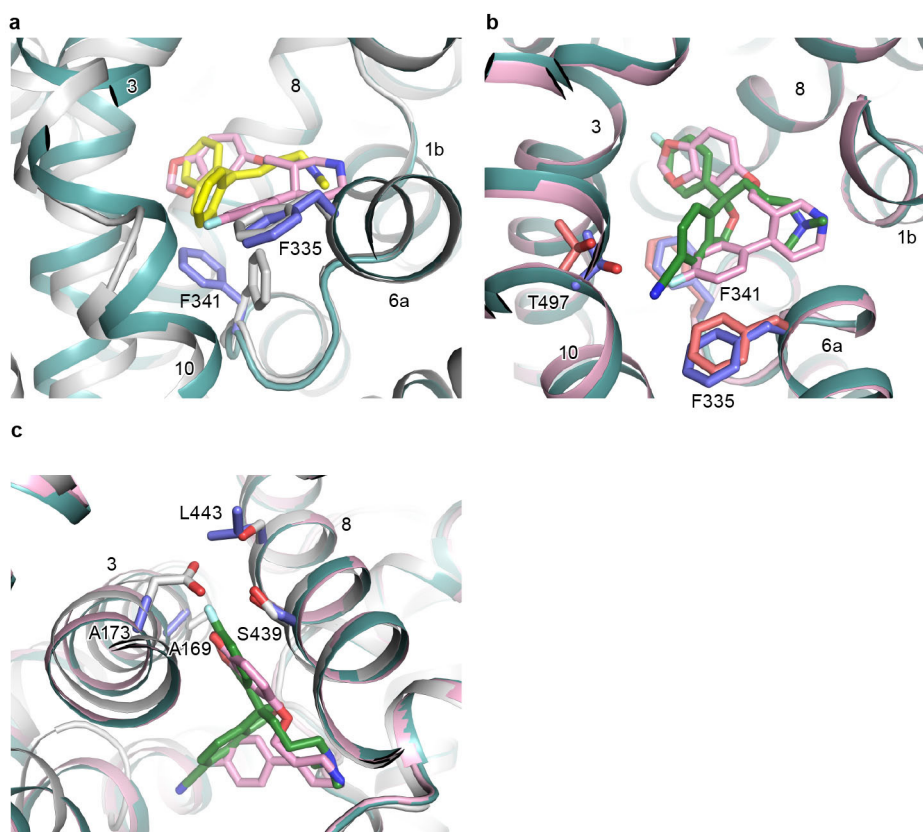
Extended Data Figure 1 | Construct design and secondary structure. Thrombin digestion sites were introduced within the N- and C-terminal regions before Gln76 and after Thr618. Mutations which were introduced to increase thermostability (Tyr110Ala, Ile291Ala and Thr439Ser) are indicated (red star). Surface exposed cysteine residues were mutated to alanine (Cys554 and Cys580) and indicated by a blue star. Residues that have no electron density are boxed in green. Secondary structure was analysed using DSSP (<http://swift.cmbi.ru.nl/gv/dssp/>) and displayed using ENDScript (<http://endscript.ibcp.fr/>). Secondary structure elements are shown using the following symbols: α -helix (α), β -strand (β), π -helix (π), 3_{10} helix (η), β -turn (TT letters), α -turn (TTT letters).

Locations of carbohydrate (red, "/) and disulfide bonding cysteine (green digits) residues are also shown. A-C in italic means the residue has a crystallographic contact with a residue in chain A-C. The '#' symbol identifies a contact between two residues along the crystallographic two-fold axis of symmetry. Contacts between transporter residues and small molecules in the range of 3.2–5.0 Å are also indicated (black, "/). Hydrophathy is calculated according to Kyte and Doolittle⁵⁹ and shown with pink as hydrophobic ($H > 1.5$), cyan as hydrophilic ($H < 1.5$), and grey as intermediate. The secondary structure of the dopamine transporter (4M48) is shown for comparison.



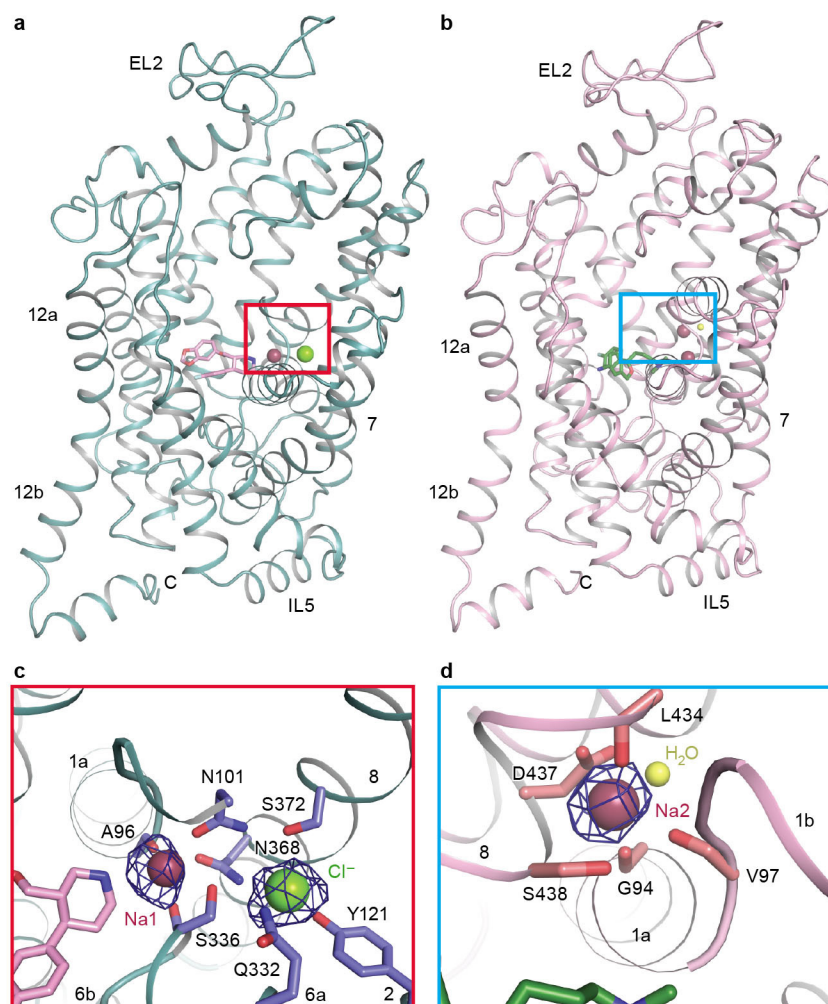
Extended Data Figure 2 | Comparison of the ts3 and ts2 structures, crystal packing and antibody structure. **a**, Superposition of the ts2 (blue) and ts3 (grey) transporters, each in complex with paroxetine using all atoms (Extended Data Table 3). Paroxetine (pink sticks) and thermostabilizing mutations (yellow spheres). **b**, Position of amino acid changes due to single nucleotide polymorphisms and mutants associated with psychiatric disorders (yellow). Paroxetine is shown in pink. **c**, SERT is shown in green, Fab heavy chain (orange), light chain (blue). SERT molecules pack into the crystal lattice with SERT–SERT interface occurring along the kink of TM12 helices related by the crystallographic two-fold axis (blue box). **d**, Rotation by 90° reveals further lattice contacts.

Red box shows interface between Fab, EL2 and EL4. We predict that this interface contains the high-affinity interaction of the Fab with EL2 and EL4. Also shown is an EL2–EL2 interaction between symmetry related molecules as well as a Fab–EL2 interface in the asymmetric unit. Purple box shows interface between Fab variable domains. Black box shows crystal contact between the C-terminal helix and the Fab constant domain. **e**, The binding site of the 8B6 Fab is made up of interactions of residues from EL2 and EL4 (sticks). **f**, Comparison of the high resolution Fab structure (grey) with SERT-bound Fab (Extended Data Table 3). The largest structural changes occur in the complementary determining regions (CDRs).



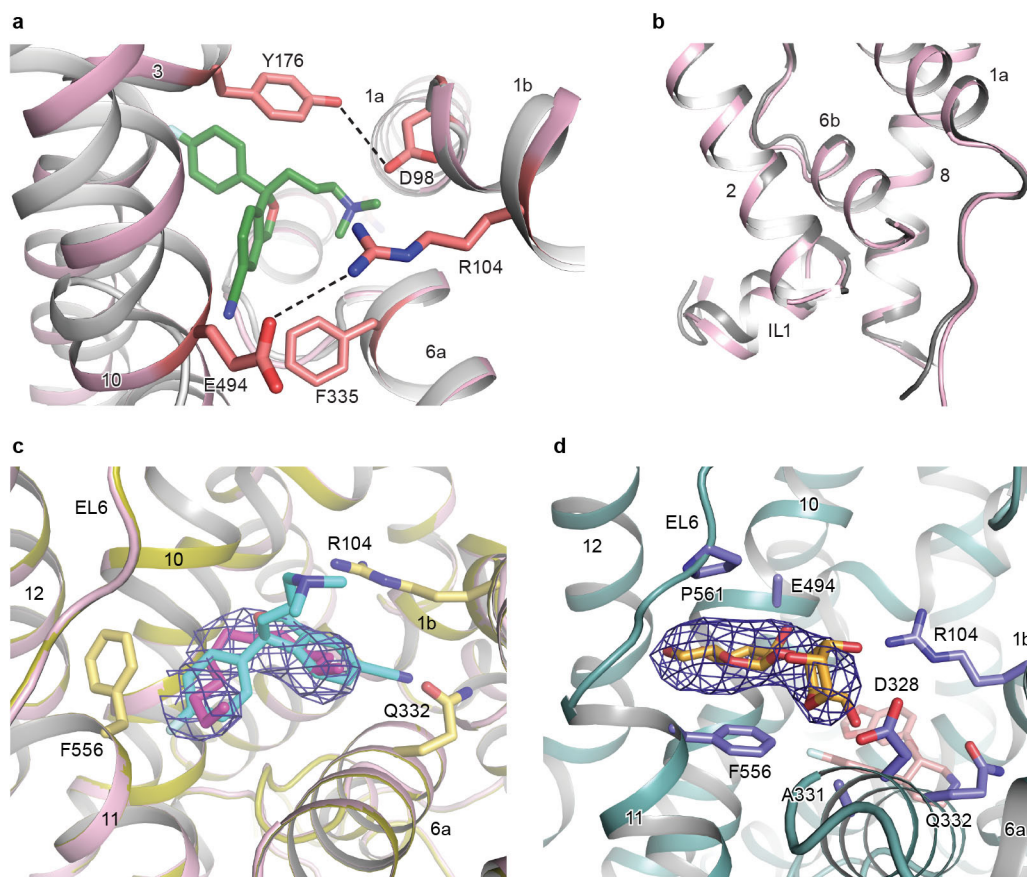
Extended Data Figure 3 | Comparison of ligand binding in SERT and in DAT. **a**, Comparison of SERT bound to paroxetine with dDAT (4M48) bound to nortriptyline (yellow); superposition based on TM1–TM12. SERT is shown in blue and DAT in grey. **b**, Alignment of paroxetine (blue) and (S)-citalopram (pink) structures using all atoms in superposition (Extended Data Table 3). Residues interacting with the antidepressant

molecules are shown as sticks. Paroxetine (pink) and (S)-citalopram (green) are shown as sticks. **c**, Insertion of benzodioxol and fluorophenyl groups of paroxetine and (S)-citalopram into a cavity in subsite B made up of Leu443, Ala169, Ala173 and Ser439. Note that Ser439 is equivalent to Thr439 in wild-type SERT. Equivalent residues in dDAT are shown in grey.



Extended Data Figure 4 | Ion-binding sites. **a**, Overall view of the Na1 and Cl⁻ ion binding sites in the paroxetine bound transporter. Na⁺ (salmon) and Cl⁻ (green) are shown as spheres. Paroxetine is shown as pink sticks. **b**, Overall view of the (S)-citalopram bound transporter showing the Na2 binding site; (S)-citalopram (green sticks). **c**, Residues

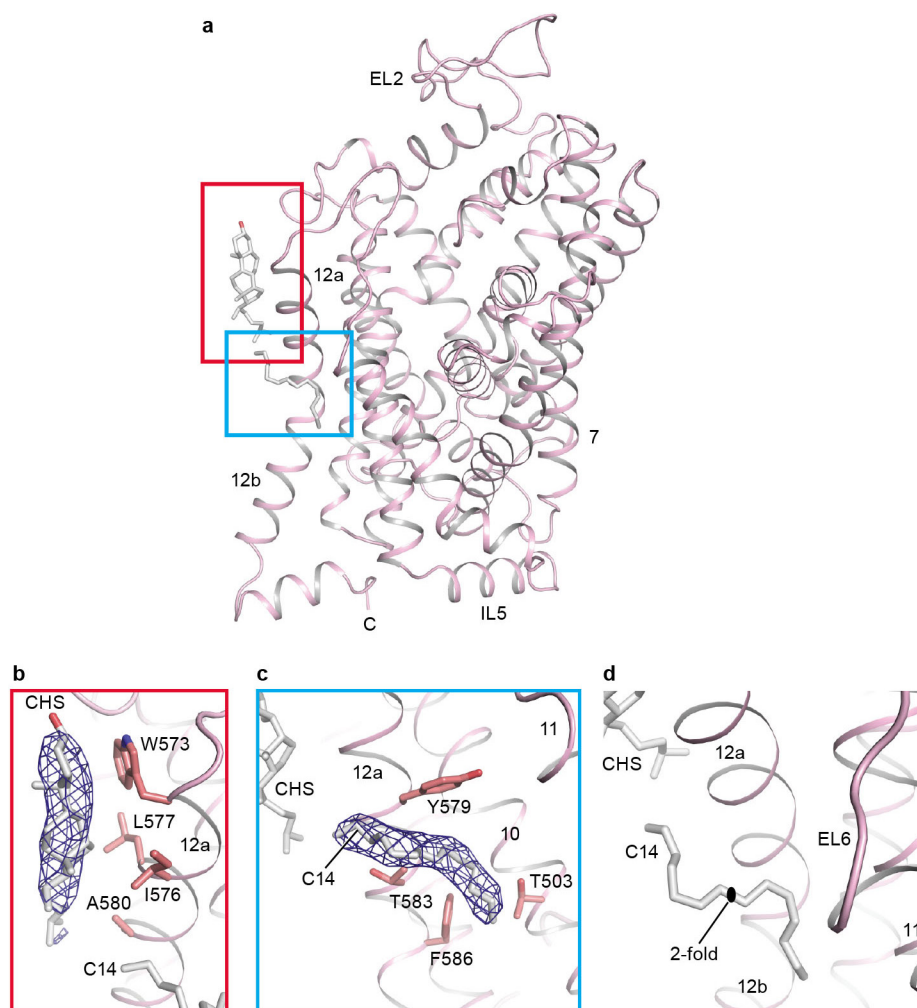
coordinating Na1 and Cl⁻. Ion $F_o - F_c$ omit densities are shown at 2σ and 3σ for Na1 and Cl⁻. **d**, Residues coordinating Na2. $F_o - F_c$ omit density is shown at 4σ . A water molecule is shown as a yellow sphere. Coordination distances are given in Extended Data Table 4.



Extended Data Figure 5 | Extracellular and intracellular gates and the allosteric site of paroxetine and partially occupied (S)-citalopram.

a, The extracellular gate of the SERT-(S)-citalopram complex is shown, with (S)-citalopram bound to the central site. The width of the gate is depicted by the distances between Tyr176 and Phe335 (10.3 Å, CD1-CE2), Asp98 and Tyr176 (4.0 Å, OD2-OH), Glu494 and Arg104 (4.9 Å, OE1-NH1) dDAT (grey) is shown for comparison. **b**, Comparison of the intracellular gate of SERT (pink) versus DAT (4M48, grey). Superpositions were made by alignment of TM1-TM12 of SERT with dDAT. **c**, The allosteric site containing fully occupied (S)-citalopram (pink) was superposed with the partially occupied structure (olive). The $F_o - F_c$ omit

density (blue mesh) of the partially occupied structure is shown at 2σ . (S)-citalopram is shown in green sticks. A 12-carbon chain (magenta) was modelled into this density but could instead represent a partially occupied (S)-citalopram. The structure with partial (S)-citalopram occupancy at the allosteric site was derived from crystals grown in the presence of 10 μ M ligand. Crystals with a higher occupancy at the allosteric site were soaked in a solution containing 5 mM (S)-citalopram before crystal cryo protection. **d**, The paroxetine-bound transporter contains a maltose detergent headgroup (orange) bound to the allosteric site. $F_o - F_c$ maltose omit density at 3σ .



Extended Data Figure 6 | Cholesteryl hemisuccinate and tetradecane binding sites. **a**, Overall view of the (S)-citalopram-bound structure showing CHS (red box) and tetradecane (C14, blue box). **b**, Zoomed view of the CHS binding site. Residues near CHS are shown as sticks. The $F_o - F_c$ omit density map is shown at 3σ . **c**, Binding of tetradecane.

The $F_o - F_c$ omit density map is contoured at 4σ . **d**, Tetradecane was modelled on a two-fold axis of symmetry with partial occupancy as a single molecule. On the basis of the density, it is unclear if this molecule represents the alkyl chain of a lipid, detergent, or a molecule of PEG.

Extended Data Table 1 | Data collection and refinement statistics

	Paroxetine ts3 *	Paroxetine ts2 †	(S)-citalopram ts3 ‡	(S)-citalopram soaked ts3 §	8B6 Fab
Data collection	APS24-IDE	ALS 5.0.2	ALS 5.0.2	ALS 5.0.2	ALS 5.0.2
Space group	C222 ₁	C222 ₁	C222 ₁	C222 ₁	P4 ₃ 2 ₁ 2
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	129.2, 162.8, 140.4	129.8, 162.8, 140.1	129.7, 163.7, 140.6	129.9, 163.2, 140.5	81.6, 81.6, 142.2
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength	0.979	1.000	1.000	0.978	1.000
Resolution (Å)	53.17-3.14 (3.25-3.14) ¶	40.50-4.53 (4.69-4.53)	53.33-3.15 (3.26-3.15)	47.68-3.24 (3.36-3.24)	57.67-1.62 (1.68-1.62)
<i>R</i> _{meas}	12.9 (63.5)	20.0 (>100)	6.7 (57.4)	8.3 (60.3)	5.7 (>100)
<i>I</i> / σ <i>I</i>	15.20 (3.1)	8.20 (0.63)	15.27 (1.99)	14.32 (2.27)	18.9 (0.61)
Completeness (%)	99.9 (100.0)	99.4 (95.5)	99.1 (98.9)	99.0 (92.0)	97.7 (76.6)
Redundancy	17.7 (12.5)	7.1 (5.9)	5.5 (4.0)	10.5 (6.6)	6.7 (2.9)
CC _{1/2} (%)	99.9 (13.8)	99.9 (24.5)	99.9 (14.0)	99.9 (11.0)	99.9 (21.5)
Refinement					
Resolution (Å)	53.17-3.14	40.50-4.53	53.33-3.15	47.68-3.24	57.67-1.62
No. reflections	26151 (2585)	8837 (828)	25995 (2577)	23870 (2178)	59809 (4634)
<i>R</i> _{work} / <i>R</i> _{free}	23.8 (38.3) / 27.0 (41.9)	28.0 (38.9) / 31.7 (37.1)	24.0 (39.1) / 27.6 (40.0)	23.5 (37.5) / 27.6 (38.0)	17.8 (42.4) / 19.8 (44.6)
No. atoms	7631	7544	7616	7633	3604
Protein	7526	7530	7515	7503	3302
Ligand/ion	104	14	100	129	N/A
Water	1	0	1	1	302
B-factors					
Protein	157.0	398.13	159.9	175.8	40.1
Ligand/ion	158.3	302.8	157.0	173.0	N/A
Water	133.5	N/A	103.3	145.0	43.0
R.m.s deviations					
Bond lengths (Å)	0.008	0.004	0.008	0.015	0.012
Bond angles (°)	0.80	0.93	1.026	1.085	1.15
Ramachandran plot					
Favored (%)	95.7	96.1	95.6	95.7	98.1
Allowed (%)	4.3	3.9	4.4	4.3	1.9
Disallowed (%)	0	0	0	0	0

*Six crystals were merged for the paroxetine ts3 structure using microdiffraction assembly.

†A single crystal was used for the paroxetine ts2 structure.

‡A single crystal was used for the (S)-citalopram ts3 structure and processed by microdiffraction assembly.

§Three crystals were merged for the (S)-citalopram-soaked ts3 structure using microdiffraction assembly.

||A single crystal was used for the 8B6 Fab structure.

¶Highest resolution shell is shown in parentheses.

5% of reflections were used for calculation of *R*_{free}.

Extended Data Table 2 | Anomalous data collection and refinement statistics

	Br-citalopram ts3 *	Br-citalopram soaked ts3 †
Data collection	APS24-IDC	ALS 5.0.2
Space group	C222 ₁	C222 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	129.6, 164.0, 140.2	129.6, 163.4, 140.5
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength	0.902	0.900
Resolution (Å)	101.7–3.40 (3.52–3.40) ‡	82.29–3.49 (3.61–3.48)
<i>R</i> _{meas}	6.3 (58.2)	8.8 (59.5)
<i>I</i> / σ <i>I</i>	17.60 (2.44)	17.76 (2.52)
Completeness (%)	99.9 (97.0)	99.5 (95.3)
Redundancy	10.7 (7.0)	11.0 (6.7)
CC _{1/2} (%)	100.0 (23.2)	99.9 (12.5)
Refinement		
Resolution (Å)	101.7–3.40	82.29–3.49
No. reflections	20919 (2004)	19231 (1812)
<i>R</i> _{work} / <i>R</i> _{free}	26.6 (39.9) / 28.5 (47.0)	24.4 (36.5) / 28.6 (36.3)
No. atoms	7610	7610
Protein	7512	7497
Ligand/ion	98	113
Water	0	0
B-factors		
Protein	220.8	175.2
Ligand/ion	200.6	151.0
Water	N/A	N/A
R.m.s deviations		
Bond lengths (Å)	0.002	0.014
Bond angles (°)	0.665	0.737
Ramachandran plot		
Favored (%)	95.8	95.5
Allowed (%)	4.2	4.5
Disallowed (%)	0	0

*A single crystal was used for the Br-citalopram ts3 structure and processed by microdiffraction assembly.

†A single crystal was used for the Br-citalopram-soaked ts3 structure and processed by microdiffraction assembly.

‡Highest resolution shell is shown in parentheses.

5% of reflections were used for calculation of *R*_{free}.

Extended Data Table 3 | Superpositions of DAT, LeuT, ts2 SERT and ts3 (S)-citalopram conformational states versus ts3 SERT (paroxetine) and comparison of high-resolution Fab versus SERT-bound Fab (paroxetine)

Structure	Rmsd (Å) / Cα #
SERT TM1-12 vs. DAT (comp. inh: nortriptyline; 4M48)	0.7 / 272
SERT TM1-12 vs. DAT (sub. analog: 3,4-dichlorophenethylamine; 4XPA)	0.8 / 283
SERT TM1,6 vs. DAT (comp. inh: nortriptyline; 4M48)	0.4 / 50
SERT TM1,6 vs. DAT (sub. analog: 3,4-dichlorophenethylamine; 4XPA)	0.6 / 52
SERT TM1-8 vs. DAT (comp. inh: nortriptyline; 4M48)	0.5 / 200
SERT TM9-12 vs. DAT (comp. inh: nortriptyline; 4M48)	1.2 / 137
SERT EL2, EL4, EL6 vs. DAT (comp. inh: nortriptyline; 4M48)	1.2 / 80
SERT TM1-12 vs. LeuT (sub. free; 3TT1)	1.3 / 260
SERT TM1-12 vs. LeuT (comp. inh: tryptophan; 3F3A)	1.4 / 249
SERT TM1-12 vs. LeuT (sub. bound: leucine; 2A65)	1.7 / 285
SERT TM1-12 vs. LeuT (inward-open; 3TT3)	2.9 / 282
ts3 (S)-citalopram SERT vs. ts3 paroxetine SERT	0.3 / 489
ts2 SERT vs. ts3 SERT	0.1 / 514
High resolution Fab vs. SERT-bound Fab	0.8 / 414

Superpositions were done by overlapping the C α atoms of DAT, LeuT or SERT ts2 over C α atoms of SERT ts3 (paroxetine) using PyMOL. A cutoff of 2.0 Å RMS was used throughout five cycles of superposition to reject outliers.

Extended Data Table 4 | Ion-binding sites and coordination distances

Site	Coordinating group	Location	Distance* (Å)
Na1 †	Ala 96-CO-	TM1a	2.3
	Asn 101 Oδ1	TM1b	2.3
	Ser 336 Oy	TM6a	2.5
	Ser 336-CO	TM6a	2.4
	Asn 368 Oδ1	TM7	2.6
	mean distance		2.4
Cl ⁻ †	Tyr 121 OH	TM2	2.6
	Gln 332 Nε	TM6a	3.1
	Ser 336 Oy	TM6a	3.6
	Ser 372 Oy	TM7	3.0
	mean distance		3.1
Na2 ‡	Gly 94-CO-	TM1a	2.5
	Val 97-CO-	TM1b	2.4
	Leu 434-CO-	TM8	2.3
	Asp 437 Oδ1	TM8	2.4
	Ser 438 Oy	TM8	2.5
	mean distance		2.4

*Interatomic distance between coordinating atom of protein chain to the bound ion(s).

†Determined from the ts3 paroxetine-bound structure.

‡Determined from the ts3 (S)-citalopram-bound structure.

A 17-billion-solar-mass black hole in a group galaxy with a diffuse core

Jens Thomas^{1,2}, Chung-Pei Ma³, Nicholas J. McConnell⁴, Jenny E. Greene⁵, John P. Blakeslee⁴ & Ryan Janish⁶

Quasars are associated with and powered by the accretion of material onto massive black holes; the detection of highly luminous quasars with redshifts greater than $z=6$ suggests that black holes of up to ten billion solar masses already existed 13 billion years ago¹. Two possible present-day ‘dormant’ descendants of this population of ‘active’ black holes have been found² in the galaxies NGC 3842 and NGC 4889 at the centres of the Leo and Coma galaxy clusters, which together form the central region of the Great Wall³—the largest local structure of galaxies. The most luminous quasars, however, are not confined to such high-density regions of the early Universe^{4,5}; yet dormant black holes of this high mass have not yet been found outside of modern-day rich clusters. Here we report observations of the stellar velocity distribution in the galaxy NGC 1600—a relatively isolated elliptical galaxy near the centre of a galaxy group at a distance of 64 megaparsecs from Earth. We use orbit superposition models to determine that the black hole at the centre of NGC 1600 has a mass of 17 billion solar masses. The spatial distribution of stars near the centre of NGC 1600 is rather diffuse. We find that the region of depleted stellar density in the cores of massive elliptical galaxies extends over the same radius as the gravitational sphere of influence of the central black holes, and interpret this as the dynamical imprint of the black holes.

We observed NGC 1600 (Fig. 1) as part of the MASSIVE Survey⁶, the aim of which is to study the structure, dynamics, and formation history of the 100 most massive early-type galaxies within 108 megaparsecs (Mpc) of Earth. This volume-limited survey probes galaxies with stellar masses above $5 \times 10^{11} M_{\odot}$ (where M_{\odot} is the mass of the Sun) in diverse, large-scale environments that have not been systematically studied before. The stellar mass ($8.3 \times 10^{11} M_{\odot}$), halo mass ($\sim 1.5 \times 10^{14} M_{\odot}$), and distance (64 Mpc) of NGC 1600 are fairly typical of the galaxies in the survey. We obtained stellar spectra covering the central 5-arcsec-by-7-arcsec region of NGC 1600 with roughly 0.6-arcsec spatial resolution, using the Gemini multi-object spectrograph (GMOS) at the Gemini North Telescope. We further obtained large-area (107-arcsec-by-107-arcsec) stellar spectra of NGC 1600 using the Mitchell integral field spectrograph (IFS) at the McDonald Observatory. The stellar luminosity distribution of the galaxy is provided by surface photometry from the Hubble Space Telescope (HST) and the Kitt Peak National Observatory⁷.

We measured the distribution of the line-of-sight stellar velocities at 86 locations in NGC 1600 by modelling the deep calcium triplet absorption lines in our GMOS IFS spectra and several optical absorption features in our Mitchell IFS spectra. The galaxy shows little rotation (less than 30 km s^{-1}), and the line-of-sight velocity dispersion rises from $235\text{--}275 \text{ km s}^{-1}$ at large radii to a maximum value of 359 km s^{-1} near the centre, consistent with previous long-slit measurements⁸.

We used orbit superposition models⁹ to determine the mass of the central black hole, M_{BH} , of NGC 1600: we find a value for M_{BH} of $(1.7 \pm 0.15) \times 10^{10} M_{\odot}$ (68% confidence interval). We rule out statistically the possibility that the M_{BH} is less than $10^{10} M_{\odot}$ with more than 99.9%

confidence (see Methods). Defining the sphere of influence of the black hole, r_{SOI} , as the radius at which the enclosed stellar mass equals M_{BH} , we find $r_{\text{SOI}} = 3.8 \text{ arcsec}$ (or 1.2 kiloparsec, kpc) for NGC 1600, using our measured M_{BH} and our calculated value for M_{\star}/L of $(4.0 \pm 0.15) M_{\odot}/L_{\odot}$ (in the R-band; L_{\odot} is the solar luminosity). Our velocity data resolve the central spatial region of NGC 1600 down to about 200 pc, where M_{BH} exceeds the enclosed stellar mass by a factor of 100. Even if the unresolved stellar mass near the centre were ten times larger because of, for example, an extreme population of undetected dwarf stars, this would not have a measurable effect on M_{BH} (see Methods).

The stellar bulge mass of NGC 1600 is $8.3 \times 10^{11} M_{\odot}$, according to our dynamical modelling, and is consistent with the value inferred from the absolute K-band magnitude of -25.99 for NGC 1600 (ref. 6). The M_{BH} of NGC 1600 is 2.1 per cent of its bulge mass (M_{bulge})—three to nine times more than the percentage predicted from the known scaling relations of black-hole and galaxy bulge mass^{10–12}. Other galaxies with high M_{BH} -to- M_{bulge} ratios are all notable for their compact size, suggesting that they are tidally stripped objects or relics from the young Universe with stunted late-time growth. In contrast, NGC 1600 is much less compact, as judged by its half-light radius, which is comparable to that of other giant elliptical galaxies of similar luminosity. The M_{BH} of NGC 1600 is also about ten times larger than the mass expected from its average stellar velocity dispersion (Extended Data Fig. 1).

NGC 1600 has a remarkably faint and flat core: it has the largest core radius detected in a study of 219 early-type galaxies using HST

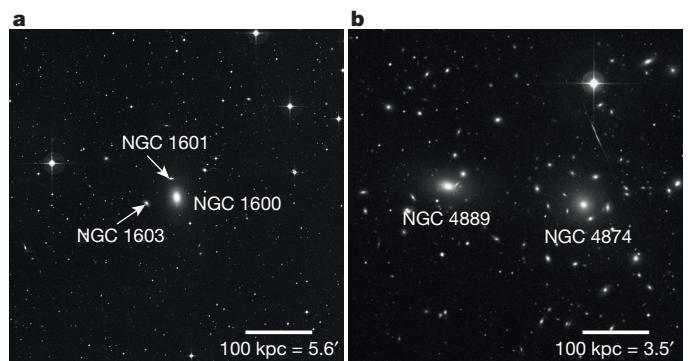


Figure 1 | Environment of NGC 1600 versus that of NGC 4889. **a**, The central $500 \text{ kpc} \times 500 \text{ kpc}$ of the NGC 1600 group of galaxies; this group has a total mass^{24,25} of about $1.5 \times 10^{14} M_{\odot}$ and an X-ray luminosity²⁶ of $3 \times 10^{41} \text{ erg s}^{-1}$. The two closest companion galaxies of NGC 1600 (NGC 1601 and NGC 1603) are nearly eight times fainter than NGC 1600. **b**, The innermost $500 \text{ kpc} \times 500 \text{ kpc}$ of the Coma Cluster, which contains more than 1,000 known galaxies, and is at least ten times more massive^{27,28} than the NGC 1600 group, and 1,000 times more X-ray-luminous²⁹. NGC 4889 has twice the stellar mass of NGC 1600 and a nearly equally luminous neighbour galaxy (NGC 4874). Both images are from the Second Palomar Observatory Sky Survey (R-band; north is at the top, and east to the left).

¹Max Planck-Institute for Extraterrestrial Physics, Giessenbachstraße 1, D-85741 Garching, Germany. ²Universitätssternwarte München, Scheinerstraße 1, D-81679 München, Germany.

³Department of Astronomy, University of California, Berkeley, California 94720, USA. ⁴Dominion Astrophysical Observatory, NRC Herzberg Institute of Astrophysics, Victoria, British Columbia V9E 2E7, Canada. ⁵Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA. ⁶Department of Physics, University of California, Berkeley, California 94720, USA.

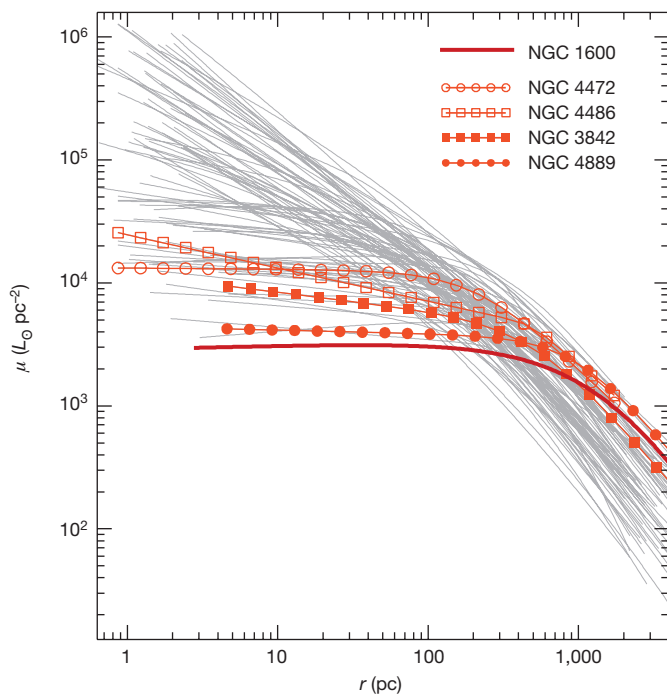


Figure 2 | Central stellar light profiles for NGC 1600 and for a sample of other core and coreless elliptical galaxies. Surface brightness profiles (in the V-band) are shown for a sample of galaxies, on the basis of HST observations¹³ up to a distance of 100 Mpc from Earth. NGC 4889, at 102 Mpc, is included because its black-hole mass has been measured². μ is the surface brightness and r is the galactic radius at which the brightness was measured. Lower-luminosity elliptical galaxies typically have rising light profiles towards the galactic centres (steep grey curves), whereas NGC 1600 and other very massive elliptical galaxies often exhibit a marked deficit of stars in the central region (red curves). Highlighted are the brightest galaxies in the Leo Cluster (NGC 3842) and the Coma Cluster (NGC 4889), and the brightest (NGC 4472) and central (NGC 4486 or M87) galaxies of the Virgo Cluster. The stellar core in NGC 1600 (dark red curve) is the faintest known among all galaxies for which dynamical M_{BH} measurements are available.

observations¹³, and a lower central surface brightness than any other galaxy within 100 Mpc of Earth in this sample ($3,100 L_{\odot} \text{pc}^{-2}$; Fig. 2). Such a diffuse light distribution indicates a substantial deficit of stars in the central region of NGC 1600 in comparison with lower-luminosity elliptical galaxies, which typically show rising light profiles towards the galactic centres down to the smallest observable radius. Like NGC 1600, other very massive elliptical galaxies often exhibit a cored light profile, where the steep light distribution characteristic of the outer part of the galaxy flattens to a nearly constant surface brightness at small radius (Fig. 2). A plausible mechanism for creating depleted stellar cores is via three-body gravitational slingshots that scatter stars passing close to a supermassive black-hole binary to larger radii. While the stars are being scattered, the orbit of the black-hole binary shrinks and the binary will emit gravitational waves if it coalesces¹⁴.

In Fig. 3 we present a new correlation between the radius, r_b , of the galaxy core in the observed light profile and the radius of the black hole's sphere of influence, for NGC 1600 and for a sample of 20 other core galaxies with reliable M_{BH} measurements¹⁵. We find a tight connection between the two radii, and the best-fit relation is consistent with $r_b = r_{\text{SOI}}$. The intrinsic scatter of 0.17 dex in the $r_b - r_{\text{SOI}}$ relation is a factor of two smaller than that in the known scaling relations between black-hole mass and galaxy properties. This small scatter holds over a wide range of galaxy environments sampled in the 21 core galaxies. Our finding that the two radii, r_b and r_{SOI} , are statistically indistinguishable with a small scatter strongly indicates that the core-formation mechanism is homogeneous and is closely connected to the central black

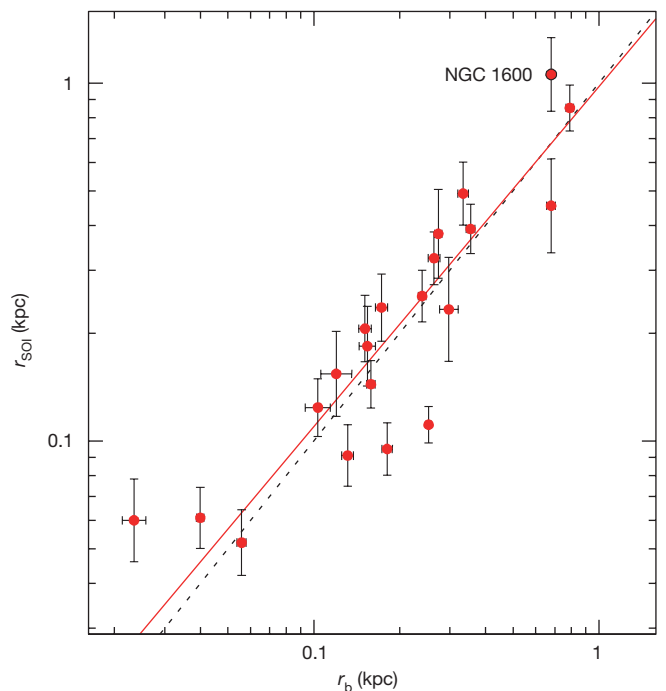


Figure 3 | Black-hole sphere-of-influence radii and galaxy core radii.

The radius of the black-hole sphere-of-influence, r_{SOI} , of NGC 1600 and 20 core galaxies with dynamical M_{BH} measurements¹⁵ is plotted against the core radius, r_b , of each host galaxy (with 1σ error bars). We calculate r_{SOI} directly from the measured M_{BH} and the de-projected stellar density profile of each galaxy. We calculate the core radius from fits to a core-Sérsic function³⁰. Our best-fit linear correlation is $\log_{10}(r_{\text{SOI}}/\text{kpc}) = (-0.01 \pm 0.29) + (0.95 \pm 0.08)\log_{10}(r_b/\text{kpc})$ (straight red line), which is statistically consistent with $r_b = r_{\text{SOI}}$ (dotted black line). The intrinsic scatter of this relation is $\epsilon = 0.17 \pm 0.04$.

hole. Gravitational core scouring by black-hole binaries can consistently produce the observed homology in the light profiles of galaxy cores^{16,17} and the velocity anisotropy in stellar orbits¹⁸ (see Methods).

Figure 4 shows the correlation between the core radius, r_b , and the measured M_{BH} for NGC 1600 and the same sample of galaxies as is shown in Fig. 3. We find an intrinsic scatter of 0.3 dex in the $M_{\text{BH}} - r_b$ relation for these 21 core galaxies, in comparison to a root-mean-squared scatter of 0.41–0.44 dex in the $M_{\text{BH}} - \sigma$ relation (where σ is the stellar velocity dispersion) for the same sample (Extended Data Fig. 1). The M_{BH} values in core galaxies generally do not correlate well with the stellar velocity dispersion^{10,11,19}. Simulated mergers of elliptical galaxies also suggest that the $M_{\text{BH}} - \sigma$ correlation may steepen or disappear altogether at the high-mass end²⁰. The most massive black holes discovered thus far predominantly reside at the centres of massive galaxies containing stellar cores. For this high-mass regime, our findings indicate that the core radius of the host galaxy is more robust than its velocity dispersion as a proxy for M_{BH} .

The Schwarzschild radius (or ‘event horizon’) of the black hole in NGC 1600 is $5 \times 10^{10} \text{ km}$, or 335 AU, subtending an angle of $5.3 \mu\text{as}$ on the sky (at a distance of 64 Mpc). Recent mass measurements^{21,22} of the black hole in the galaxy M87 range from $3.3 \times 10^9 M_{\odot}$ to $6.2 \times 10^9 M_{\odot}$, corresponding to a Schwarzschild radius of $3.8\text{--}7.3 \mu\text{as}$ (at 16.7 Mpc). For comparison, the Schwarzschild radius of Sagittarius A* is $10 \mu\text{as}$. Thus, after the Milky Way and possibly M87, NGC 1600 contains the next most easily resolvable black hole, and is a candidate for observations with the Event Horizon Telescope²³.

Black holes with masses of about $10^{10} M_{\odot}$ are observed as quasars in the young Universe¹. Finding the dormant-black-hole descendants of these luminous quasars and understanding their ancestral lineages have been strong motivations for the search for very massive black holes nearby. At redshifts of 2 to 3 (about 10 billion years ago), when the

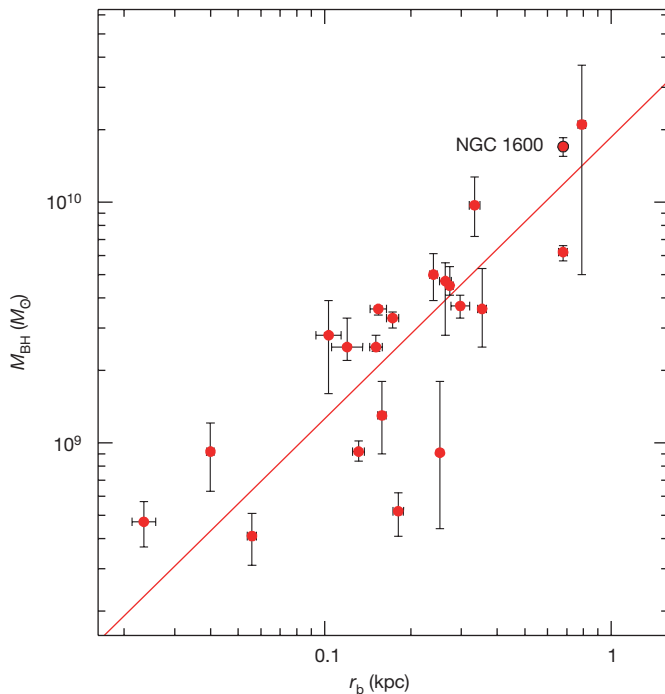


Figure 4 | Black-hole mass and galaxy core radius. The black-hole mass, M_{BH} , is plotted against the core radius, r_b , for the same sample of galaxies as in Fig. 3 (and with 1σ error bars). The straight line shows our best fit for a constant-exponent power-law relation between M_{BH} and r_b : $\log_{10}(M_{\text{BH}}/M_{\odot}) = (10.27 \pm 0.51) + (1.17 \pm 0.14)\log_{10}(r_b/\text{kpc})$. The intrinsic scatter of this relation is $\epsilon = 0.29 \pm 0.07$.

quasar activity peaks, the luminosities of the brightest quasars (more than $10^{48} \text{ erg s}^{-1}$) suggest the existence of very massive black holes (with masses of more than $5 \times 10^9 M_{\odot}$). Yet, on average, these quasars do not appear to inhabit environments that are notably different from those of their less luminous counterparts^{4,5}. The 17-billion-solar-mass black hole in NGC 1600 is a possible descendant of a luminous quasar in an environment outside the richest structures of galaxies. Very massive dark-matter haloes are therefore not necessary for growing very massive black holes and their luminous quasar progenitors. The number density of dark-matter haloes of comparable mass to that of NGC 1600 is about $2 \times 10^{-5} \text{ Mpc}^{-3}$, making such haloes about 50 times more abundant than rich structures like the Coma Cluster. At present, we do not know whether dormant black holes with masses greater than $10^{10} M_{\odot}$ are common in other nearby $\sim 10^{14} M_{\odot}$ haloes with massive galaxies, or whether other properties of the NGC 1600 group—such as the large magnitude gap, fossil-group-like evolved dynamical state²⁴, and isolation—are necessary ingredients for cultivating such massive black holes. Our ongoing observations of massive galaxies⁶ will soon reveal whether the extreme black hole in NGC 1600 is a rare find in an unusual environment, or the tip of an iceberg.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 October 2015; accepted 29 January 2016.

Published online 6 April 2016.

- Wu, X.-B. *et al.* An ultraluminous quasar with a twelve-billion-solar-mass black hole at redshift 6.30. *Nature* **518**, 512–515 (2015).
- McConnell, N. J. *et al.* Two ten-billion-solar-mass black holes at the centres of giant elliptical galaxies. *Nature* **480**, 215–218 (2011).
- Geller, M. J. & Huchra, J. P. Mapping the universe. *Science* **246**, 897–903 (1989).
- Trainor, R. F. & Steidel, C. C. The halo masses and galaxy environments of hyperluminous QSOs at $z \approx 2.7$ in the Keck baryonic structure survey. *Astrophys. J.* **752**, 39 (2012).
- Fanidakis, N., Macciò, A. V., Baugh, C. M., Lacey, C. G. & Frenk, C. S. The most luminous quasars do not live in the most massive dark matter haloes at any redshift. *Mon. Not. R. Astron. Soc.* **436**, 315–326 (2013).

- Ma, C.-P. *et al.* The MASSIVE survey. I. A volume-limited integral-field spectroscopic study of the most massive early-type galaxies within 108 Mpc. *Astrophys. J.* **795**, 158 (2014).
- Peletier, R. F., Davies, R. L., Illingworth, G. D., Davis, L. E. & Cawson, M. CCD surface photometry of galaxies with dynamical data. II—UBR photometry of 39 elliptical galaxies. *Astron. J.* **100**, 1091–1142 (1990).
- Bender, R., Saglia, R. P. & Gerhard, O. E. Line-of-sight velocity distributions of elliptical galaxies. *Mon. Not. R. Astron. Soc.* **269**, 785–813 (1994).
- Thomas, J. *et al.* Mapping stationary axisymmetric phase-space distribution functions by orbit libraries. *Mon. Not. R. Astron. Soc.* **353**, 391–404 (2004).
- McConnell, N. J. & Ma, C.-P. Revisiting the scaling relations of black hole masses and host galaxy properties. *Astrophys. J.* **764**, 184 (2013).
- Kormendy, J. & Ho, L. C. Coevolution (or not) of supermassive black holes and host galaxies. *Ann. Rev. Astron. Astrophys.* **51**, 511–653 (2013).
- Saglia, R. P. *et al.* The SINFONI black hole survey: the black hole fundamental plane revisited and the paths of (co)evolution of supermassive black holes and bulges. *Astrophys. J.* **818**, 47 (2016).
- Lauer, T. R. *et al.* The centers of early-type galaxies with Hubble Space Telescope. VI. Bimodal central surface brightness profiles. *Astrophys. J.* **664**, 226–256 (2007).
- Begelman, M. C., Blandford, R. D. & Rees, M. J. Massive black hole binaries in active galactic nuclei. *Nature* **287**, 307–309 (1980).
- Rusli, S. P. *et al.* Depleted galaxy cores and dynamical black hole masses. *Astron. J.* **146**, 160 (2013).
- Faber, S. M. *et al.* The centers of early-type galaxies with HST. IV. Central parameter relations. *Astron. J.* **114**, 1771–1796 (1997).
- Kormendy, J. & Bender, R. Correlations between supermassive black holes, velocity dispersions, and mass deficits in elliptical galaxies with cores. *Astrophys. J.* **691**, L142–L146 (2009).
- Thomas, J., Saglia, R. P., Bender, R., Erwin, P. & Fabricius, M. The dynamical Fingerprint of core scouring in massive elliptical galaxies. *Astrophys. J.* **782**, 39 (2014).
- Kormendy, J. & Bender, R. The $L \propto \sigma^8$ correlation for elliptical galaxies with cores: relation with black hole mass. *Astrophys. J.* **769**, L5 (2013).
- Boylan-Kolchin, M., Ma, C.-P. & Quataert, E. Red mergers and the assembly of massive elliptical galaxies: the fundamental plane and its projections. *Mon. Not. R. Astron. Soc.* **369**, 1081–1089 (2006).
- Gebhardt, K. *et al.* The black hole mass in M87 from Gemini/NIFS adaptive optics observations. *Astrophys. J.* **729**, 119 (2011).
- Walsh, J. L., Barth, A. J., Ho, L. C. & Sarzi, M. The M87 black hole mass from gas-dynamical models of space telescope imaging spectrograph observations. *Astrophys. J.* **770**, 86 (2013).
- Doeleman, S. S. *et al.* Jet-launching structure resolved near the supermassive black hole in M87. *Science* **338**, 355–358 (2012).
- Smith, R. M., Martínez, V. J., Fernández-Soto, A., Ballesteros, F. J. & Ortiz-Gil, A. NGC 1600: cluster or field elliptical? *Astrophys. J.* **679**, 420–427 (2008).
- Crook, A. C. *et al.* Groups of galaxies in the two micron all sky redshift survey. *Astrophys. J.* **655**, 790–813 (2007).
- Sivakoff, G. R., Sarazin, C. L. & Carlini, J. L. Chandra observations of diffuse gas and luminous X-ray sources around the X-ray-bright elliptical galaxy NGC 1600. *Astrophys. J.* **617**, 262–280 (2004).
- Rines, K., Geller, M. J., Kurtz, M. J. & Diaferio, A. CAIRNS: the cluster and infall region nearby survey. I. Redshifts and mass profiles. *Astron. J.* **126**, 2152–2170 (2003).
- Kubo, J. M. *et al.* The mass of the Coma cluster from weak lensing in the Sloan Digital Sky Survey. *Astrophys. J.* **671**, 1466–1470 (2007).
- Ikebe, Y., Reiprich, T. H., Böhringer, H., Tanaka, Y. & Kitayama, T. A new measurement of the X-ray temperature function of clusters of galaxies. *Astron. Astrophys.* **383**, 773–790 (2002).
- Graham, A. W., Erwin, P., Trujillo, I. & Asensio Ramos, A. A new empirical model for the structural analysis of early-type galaxies, and a critical review of the Nuker model. *Astron. J.* **125**, 2951–2963 (2003).

Acknowledgements C.-P.M., J.E.G. and R.J. are supported by the National Science Foundation (NSF). J.E.G. is supported by the Miller Institute for Basic Research in Science, University of California, Berkeley. N.J.M. is supported by the Beatrice Watson Parrent Fellowship and Plaskett Fellowship. The spectroscopic data presented here were obtained from the Gemini Observatory and the McDonald Observatory. Gemini is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership. The McDonald Observatory is operated by the University of Texas at Austin. The photometric data presented here are based partly on observations made with the NASA/ESA Hubble Space Telescope, and obtained from the Hubble Legacy Archive, which is a collaboration between the Space Telescope Science Institute (STScI/NASA), the Space Telescope European Coordinating Facility (ST-ECF/ESA) and the Canadian Astronomy Data Centre (CAD/C/NRC/CSA).

Author Contributions J.T. developed and carried out the stellar orbit modelling. J.T. and C.-P.M. wrote the manuscript. C.-P.M. led the Gemini observation proposal. J.E.G. performed the stellar population analysis. N.J.M. and R.J. reduced the spectroscopic data. J.P.B. provided photometric analysis. All authors contributed to the MASSIVE Survey, the kinematic extractions, the interpretive analysis of the observations and the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.T. (jthomas@mpe.mpg.de) or C.-P.M. (cpma@berkeley.edu).

METHODS

In the first section below, we describe spectroscopic data for NGC 1600 and our procedures for measuring stellar kinematics. In the second section, we describe photometric data and the surface brightness profile of NGC 1600. In the third section, we describe our stellar orbit modelling procedures. In the fourth section we compare the stellar mass-to-light ratio obtained from the dynamical models with independent constraints from a stellar population analysis. The implied distribution of stellar orbits in NGC 1600 is described in the fifth section. Finally, in the last section we present the black-hole scaling relations for core radii, determined from an alternative fit to the light profiles.

Spectroscopic data and stellar kinematics. We obtained high-spatial-resolution spectroscopic data from GMOS-N, an IFS on the 8-metre Gemini North Telescope. We observed the central region of NGC 1600 with GMOS-N, which provides continuous two-dimensional coverage of a $5 \text{ arcsec} \times 7 \text{ arcsec}$ science field and simultaneously covers a sky field offset by 60 arcsec . Our spectra were centred on the triplet of calcium absorption lines from $8,480 \text{ \AA}$ to $8,680 \text{ \AA}$, a well studied region frequently used for stellar kinematic measurements^{31,32}. We obtained nine 1,230-second exposures of NGC 1600 over three nights of queue-mode observations in November 2014.

GMOS-N is seeing-limited, with spatial sampling of 0.2 arcsec on the IFS. We estimated the point-spread function (PSF) on each observing night by measuring the width of foreground stars in the acquisition images of NGC 1600. Our average PSF for GMOS-N has a full width at half-maximum (FWHM) of 0.6 arcsec . A Gaussian model of the PSF is included in our orbit superposition models.

We used the image reduction and analysis facility (IRAF) software package supplied by the Gemini Observatory to flat-field and wavelength-calibrate the GMOS-N data, and to extract a one-dimensional spectrum for each IFS lenslet. We developed custom routines to construct collapsed images of the galaxy and record the position of each one-dimensional spectrum with respect to the galaxy centre. The spectra were then resampled on a two-dimensional grid and binned to a consistent signal-to-noise ratio (of about 100 per pixel) using Voronoi tessellation³³. Our binning implementation imposed symmetry over four projected quadrants of the galaxy, so that the kinematic measurements could be folded into a single quadrant before orbit modelling.

We obtained wide-field spectroscopic data from the Mitchell Spectrograph³⁴ on the 2.7-metre Harlan J. Smith Telescope at the McDonald Observatory. The Mitchell Spectrograph is an optical IFS with a $107 \text{ arcsec} \times 107 \text{ arcsec}$ field of view and 246 fibres, each of 4.1 arcsec diameter. The low-resolution blue setting ($R \approx 850$) was used, providing wavelength coverage from $3,650 \text{ \AA}$ to $5,850 \text{ \AA}$, including the Ca H+K region, the G-band region, H β , the Mg b region, and several Fe absorption features. The spectral resolution varied spatially and with wavelength, with an average of 5 \AA FWHM, corresponding to a dispersion of $\sim 1.1 \text{ \AA pixel}^{-1}$ and $\sigma_{\text{inst}} \approx 100 \text{ km s}^{-1}$ in the red and $\sigma_{\text{inst}} \approx 150 \text{ km s}^{-1}$ in the blue part of the spectrum (where σ_{inst} is the instrumental resolution). Data reduction was performed using the Vaccine package³⁵. We fit Mitchell spectra with the MILES library of 985 stellar spectra³⁶ and determined the best-fit line-of-sight velocity distribution (LOSVD) for each of the 58 spatial bins^{6,37}. Our GMOS and Mitchell data are the first IFS observations of NGC 1600 and reveal a stellar velocity distribution that is well aligned with the galaxy's light distribution, indicating that NGC 1600 is axisymmetric.

Photometric data and core-Sérsic fit. We used a high-resolution image taken with the near-infrared camera and multi-object spectrometer (NICMOS) instrument on the HST to measure the central light distribution of NGC 1600. The observation (from General Observer Program number 7886) consisted of four dithered exposures of NGC 1600 taken with NICMOS camera 2 in the F160W bandpass. We downloaded the calibrated, combined image from the Hubble Legacy Archive. The image had a pixel scale of 0.05 arcsec per pixel and total exposure time of 460 seconds.

We combined the HST observations with ground-based photometric data at large radii taken from the literature⁷. The NICMOS data were calibrated to the R-band of the ground-based data by minimizing the squared magnitude differences between the two surface brightness profiles in the radial region where both data sets overlap and PSF effects are negligible ($r = 2\text{--}10 \text{ arcsec}$). Single one-dimensional profiles of the surface brightness, the ellipticity and the isophotal shape parameters³⁸ a_4 and a_6 were then constructed by using the NICMOS data at $r < 8 \text{ arcsec}$ and the ground-based data at $r \geq 8 \text{ arcsec}$. The position angle (PA) of the isophotes is constant with radius⁷ ($\Delta \text{PA} < 2^\circ$), consistent with an axisymmetric stellar distribution.

The resulting circularized surface brightness distribution of NGC 1600 is well described by a core-Sérsic function with a core radius of $r_b = 2.15 \text{ arcsec}$ (Extended Data Fig. 2). Over several orders of magnitude in radius, the light profiles of lower-luminosity elliptical galaxies follow a single Sérsic function characterized by

the Sérsic index n , the half-light radius r_e and a surface brightness scale $\mu_c = \mu(r_e)$. The core-Sérsic function combines a Sérsic profile at $r > r_b$ and a power-law distribution with slope γ at $r < r_b$. The transition is controlled by a smoothness parameter α and the surface brightness scale is $\mu_b = \mu(r_b)$. Inside $r < 5 \text{ arcsec}$, the inward extrapolation of the outer Sérsic component overpredicts the central surface brightness of NGC 1600 by about three magnitudes. From the difference between the integrated light of the Sérsic component and the actual core-Sérsic fit, we derive a 'light deficit' of $\Delta L_{\text{def}} = 9.47 \times 10^9 L_\odot$ in the centre of NGC 1600.

We use the isophotal model of NGC 1600 to compute the galaxy's intrinsic luminosity density distribution. The deprojection is nonparametric³⁹ and accounts for the observed ellipticity profile and boxy shape of NGC 1600's isophotes⁷. The same technique has been used for the dynamical modelling of other galaxies^{12,40}. **Stellar orbit models.** We generate dynamical models of NGC 1600 using Schwarzschild's orbit superposition method⁴¹. Because the two-body relaxation time of stars in massive elliptical galaxies exceeds the age of the Universe, their dynamics is governed by the collisionless Boltzmann equation. Any steady-state equilibrium solution of the Boltzmann equation can be written as a sum over single-orbit distribution functions, where the phase-space density of stars along each trajectory is constant. The total number of stars on each orbit—that is, the orbital luminosity or orbital occupation number—can take arbitrary (positive) values.

For the modelling, we assume that NGC 1600 is axisymmetric and that the stellar mass profile follows the observed light distribution with a constant stellar mass-to-light ratio, M_*/L . The models also assume a central black hole of mass M_{BH} and a cored isothermal dark-matter halo with core radius r_{DM} and asymptotic circular velocity v_{DM} . The four parameters of the mass model are constrained by the photometrically derived luminosity density and by 1,978 LOSVD data points measured from the galaxy spectra between 0.4 arcsec and $\sim 45 \text{ arcsec}$.

Given some specific values for M_*/L , M_{BH} , r_{DM} and v_{DM} , the Poisson equation is solved for the gravitational potential generated by the respective mass model and the luminosity densities, and the LOSVDs of $\sim 29,000$ representative stellar orbits are computed⁹. The orbits are convolved with the PSF of the observations and integrated over the respective areas on the sky. We use a maximum entropy technique⁴² to determine the orbital occupation numbers that minimize the χ^2 difference among the observational data and the orbit superposition model. Thousands of different mass distributions are compared to the data by systematically varying M_*/L , M_{BH} , r_{DM} and v_{DM} . We repeat the computation of the stellar orbits and the phase-space optimization independently for each model. The best-fit values and confidence intervals in M_{BH} and M_*/L are determined by evaluating the relative likelihood⁴³ for all models with different assumed values of M_{BH} , M_*/L and dark halo parameters (Extended Data Fig. 3). The same modelling technique to determine the mass of the stars, the central black hole and the dark matter halo was first applied to the central galaxy of the Virgo Cluster, M87 (ref. 44).

Previous models of NGC 1600, based on stellar velocity data only along the major and minor axes of the galaxy and with a lower spatial resolution, did not include all three mass components^{45–48}. We have tested that models without a dark-matter halo and/or without a central black hole cannot reproduce the full set of our new observations. Extended Data Fig. 4 shows our best-fit orbit model together with the velocity data for NGC 1600.

Stellar mass-to-light ratio. The stellar mass-to-light ratio derived from our dynamical modelling is $M_*/L = (4.0 \pm 0.15) M_\odot/L_\odot$. The stellar mass deficit in the core of NGC 1600 is thus $\Delta M_{\text{def}} = 3.8 \times 10^{10} M_\odot$. In other core galaxies, mass deficits have been reported to range from one to ten times the mass of the central black hole^{15,17,49}. For NGC 1600 we find a mass deficit, ΔM_{def} , of $2.2 \times M_{\text{BH}}$. Results from numerical simulations of mergers of galaxies with central black holes⁵⁰ suggest mass deficits of $\sim N \times 0.5 M_{\text{BH}}$, where N is the number of mergers.

We measure a stellar population age of $\tau \approx 10 \text{ Gyr}$, a total metallicity of $[Z/H] = 0.03$ and an iron abundance of $[\text{Fe}/H] = -0.15$ for NGC 1600, from the absorption-line strengths of hydrogen, iron and other metallicity-dependent stellar lines in the 1-kpc galaxy core. It has been reported that the fraction of low-luminosity dwarf stars in elliptical galaxies less luminous than NGC 1600 is larger than that in the Milky Way^{51–53}. Extrapolating correlations with the galaxy velocity dispersion obtained for these smaller elliptical galaxies would yield a higher stellar mass of $M_*/L \approx 6.0$ for NGC 1600. We find the galaxy-wide contribution of dwarf stars to the stellar mass in NGC 1600 to be consistent with observations of the Milky Way. Dynamical constraints on the fractional mass of dwarf stars depend on the assumed dark-matter halo profile⁵⁴ and on the assumed shape of the stellar initial mass function⁵⁵. Some nearby massive galaxies have, like NGC 1600, dwarf-star fractions consistent with that of the Milky Way⁵⁶.

The spectroscopic analysis of the age and chemical composition of the stars in NGC 1600 does not provide evidence for a notable change in the stellar population with radius. In the dynamical modelling we assume a constant M_*/L ratio throughout the galaxy. Individual massive galaxies have been reported to host

extreme populations of dwarf stars at their centres⁵⁷ that would not be detectable in our optical spectra. The respective dwarf stars can increase the central M_*/L by up to a factor of three. If the assumed constant M_*/L does not account for all of the stellar mass at the centre of NGC 1600, then the dynamical models may compensate for the missing stellar mass by overestimating M_{BH} . Extended Data Fig. 5 shows the enclosed mass distribution of NGC 1600 over the region for which we have obtained stellar velocity data. For a constant M_*/L , we find the enclosed stellar mass at the smallest observed radius ($r \approx 0.2$ kpc) to be 100 times smaller than M_{BH} . A central increase of M_*/L by a factor as extreme as ten would imply an unaccounted-for extra stellar mass of 10% of M_{BH} (dotted lines in Extended Data Fig. 5). Even in this unrealistic case, we would overestimate M_{BH} only by its one sigma measurement error. Because there is so little stellar light in the core of NGC 1600, uncertainties in the central stellar population have a negligible effect on our black-hole mass measurement.

Distribution of stellar orbits. The stars in a galaxy are collisionless and their velocity distribution can be anisotropic. We compute the intrinsic velocity dispersions of the stars along the radial and the two angular directions of a polar coordinate system— σ_r , σ_θ , and σ_ϕ , respectively—from the orbital occupation numbers of our best-fit dynamical model in 20 spherical shells centred on NGC 1600's black hole. The classic measure for the anisotropy of the stellar velocities is $\beta = 1 - \sigma_\phi^2/\sigma_r^2$, where the tangential velocity dispersion $\sigma_t = [(\sigma_\theta^2 + \sigma_\phi^2)/2]^{1/2}$ is the average of the motions in the two angular directions. Stellar orbits in core galaxies have been reported to be very uniform¹⁸. In NGC 1600 and similar galaxies with a flat central surface brightness, most of the stars inside the diffuse core region ($r < r_b$) are moving along tangential directions. With increasing distance from the centre, more and more stars are found on radially elongated orbits (Extended Data Fig. 6).

In the black-hole-binary model, the observed stellar motions are naturally explained as the leftover of the core scouring process (shaded regions in Extended Data Fig. 6). Central stars originally on radial trajectories are subject to interactions with the black-hole binary as they frequently pass the galaxy centre. Eventually, these stars get ejected to larger radii via gravitational slingshot. The stars that we observe today in the centres of core galaxies remained there because they moved (and still move) on tangential orbits that avoid the centre⁵⁸.

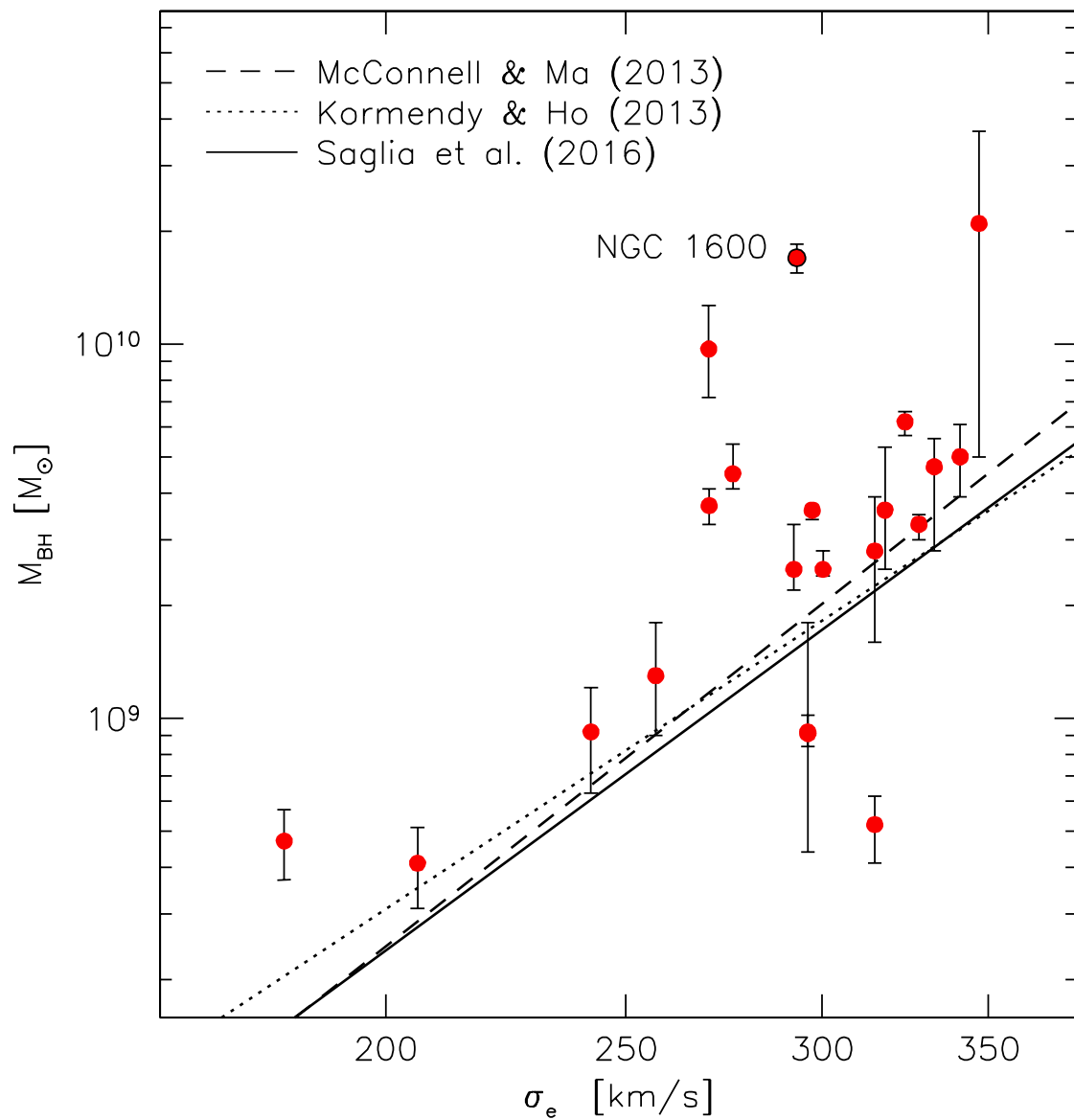
It has not yet been tested whether other black-hole activities—such as their feedback processes on ambient accreting gas—can produce the tight relations between black-hole mass, core radius, sphere of influence and mass deficit together with the observed orbital structure.

Measures of galaxy core size. While the core-Sérsic function describes galaxy light profiles from the core region out to large radii^{30,59}, the Nuker function⁶⁰ has been widely used to fit the central light profiles of galaxies observed with HST. Fifteen out of the 21 core galaxies discussed in the text also have core radii measured from Nuker fits¹³. For the Nuker r_b , we obtain $\log_{10}(r_{\text{SOI}}/\text{kpc}) = (-0.18 \pm 0.21) + (1.00 \pm 0.09)\log_{10}(r_b/\text{kpc})$ with an intrinsic scatter of $\epsilon = 0.16$, and $\log_{10}(M_{\text{BH}}/M_\odot) = (10.06 \pm 0.45) + (1.25 \pm 0.17)\log_{10}(r_b/\text{kpc})$ with an intrinsic scatter of $\epsilon = 0.31$. The Nuker r_b values were measured along the major axis of the galaxies, while the core-Sérsic r_b values discussed in the text come from the galaxies' circularized light profiles¹⁵.

Galaxy core sizes have also been quantified by the cusp radius⁶¹, r_c —that is, the radius at which the negative logarithmic slope of the surface brightness profile equals 1/2. We obtained the cusp radii of the galaxies that are shown in Figs 3 and 4 from their core-Sérsic models. Using r_c as a measure of the core size, we find $\log_{10}(r_{\text{SOI}}/\text{kpc}) = (0.06 \pm 0.28) + (0.94 \pm 0.09)\log_{10}(r_c/\text{kpc})$ (intrinsic scatter $\epsilon = 0.16$) and $\log_{10}(M_{\text{BH}}/M_\odot) = (10.37 \pm 0.60) + (1.20 \pm 0.19)\log_{10}(r_c/\text{kpc})$, $\epsilon = 0.33$. Note that the core of NGC 1550 has a slope¹⁵ of $\gamma = 0.52 \pm 0.05$. The scaling relations with r_c have been computed without including NGC 1550. The slope and the scatter in the above correlations are consistent with the results for the core-Sérsic r_b .

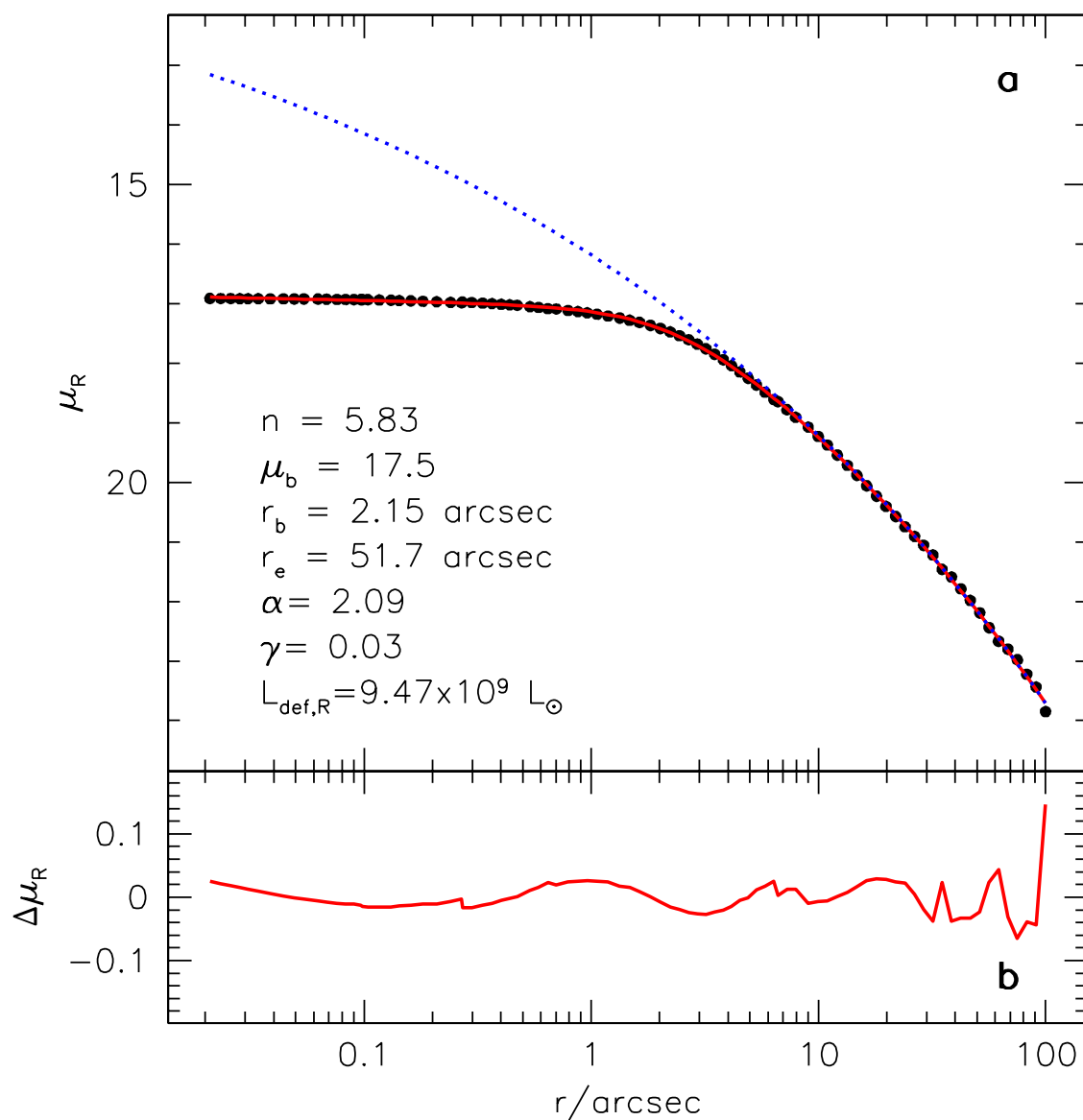
31. Barth, A. J., Ho, L. C. & Sargent, W. L. W. A study of the direct fitting method for measurement of galaxy velocity dispersions. *Astron. J.* **124**, 2607–2614 (2002).
32. McConnell, N. J. *et al.* Dynamical measurements of black hole masses in four brightest cluster galaxies at 100 Mpc. *Astrophys. J.* **756**, 179 (2012).
33. Cappellari, M. & Copin, Y. Adaptive spatial binning of integral-field spectroscopic data using Voronoi tessellations. *Mon. Not. R. Astron. Soc.* **342**, 345–354 (2003).

34. Hill, G. J., MacQueen, P. J., Palunas, P., Barnes, S. I. & Shetrone, M. D. Present and future instrumentation for the Hobby-Eberly Telescope. *Proc. SPIE* **7014**, 701406 (2008).
35. Adams, J. J. *et al.* The HETDEX pilot survey. I. Survey design, performance, and catalog of emission-line galaxies. *Astrophys. J. (Suppl.)* **192**, 5 (2011).
36. Sánchez-Blázquez, P. *et al.* Medium-resolution Isaac Newton Telescope library of empirical spectra. *Mon. Not. R. Astron. Soc.* **371**, 703–718 (2006).
37. Cappellari, M. & Emsellem, E. Parametric recovery of line-of-sight velocity distributions from absorption-line spectra of galaxies via penalized likelihood. *Pub. Astron. Soc. Pacific* **116**, 138–147 (2004).
38. Bender, R. & Moellenhoff, C. Morphological analysis of massive early-type galaxies in the Virgo cluster. *Astron. Astrophys.* **177**, 71–83 (1987).
39. Magorrian, J. Kinematical signatures of hidden stellar discs. *Mon. Not. R. Astron. Soc.* **302**, 530–536 (1999).
40. Thomas, J. *et al.* Regularized orbit models unveiling the stellar structure and dark matter halo of the Coma elliptical NGC 4807. *Mon. Not. R. Astron. Soc.* **360**, 1355–1372 (2005).
41. Schwarzschild, M. A numerical model for a triaxial stellar system in dynamical equilibrium. *Astrophys. J.* **232**, 236–247 (1979).
42. Richstone, D. O. & Tremaine, S. Maximum-entropy models of galaxies. *Astrophys. J.* **327**, 82–88 (1988).
43. McConnell, N. J. *et al.* The black hole mass in the brightest cluster galaxy NGC 6086. *Astrophys. J.* **728**, 100 (2011).
44. Gebhardt, K. & Thomas, J. The black hole mass, stellar mass-to-light ratio, and dark halo in M87. *Astrophys. J.* **700**, 1690–1701 (2009).
45. van der Marel, R. P. The velocity dispersion anisotropy and mass-to-light ratio of elliptical galaxies. *Mon. Not. R. Astron. Soc.* **253**, 710–726 (1991).
46. Magorrian, J. *et al.* The demography of massive dark objects in galaxy centers. *Astron. J.* **115**, 2285–2305 (1998).
47. Matthias, M. & Gerhard, O. Dynamics of the boxy elliptical galaxy NGC 1600. *Mon. Not. R. Astron. Soc.* **310**, 879–891 (1999).
48. Pu, S. B. *et al.* Radially extended kinematics and stellar populations of the massive ellipticals NGC 1600, NGC 4125, and NGC 7619. Constraints on the outer dark halo density profile. *Astron. Astrophys.* **516**, A4 (2010).
49. Dullo, B. T. & Graham, A. W. Depleted cores, multicomponent fits, and structural parameter relations for luminous early-type galaxies. *Mon. Not. R. Astron. Soc.* **444**, 2700–2722 (2014).
50. Merritt, D. Mass deficits, stalling radii, and the merger histories of elliptical galaxies. *Astrophys. J.* **648**, 976–986 (2006).
51. Treu, T. *et al.* The initial mass function of early-type galaxies. *Astrophys. J.* **709**, 1195–1202 (2010).
52. Conroy, C. & van Dokkum, P. G. The stellar initial mass function in early-type galaxies from absorption line spectroscopy. II. Results. *Astrophys. J.* **760**, 71 (2012).
53. Cappellari, M. *et al.* The ATLAS3D project—XX. Mass-size and mass- σ distributions of early-type galaxies: bulge fraction drives kinematics, mass-to-light ratio, molecular gas fraction and stellar initial mass function. *Mon. Not. R. Astron. Soc.* **432**, 1862–1893 (2013).
54. Thomas, J. *et al.* Dynamical masses of early-type galaxies: a comparison to lensing results and implications for the stellar initial mass function and the distribution of dark matter. *Mon. Not. R. Astron. Soc.* **415**, 545–562 (2011).
55. La Barbera, F. *et al.* SPIDER VIII—constraints on the stellar initial mass function of early-type galaxies from a variety of spectral features. *Mon. Not. R. Astron. Soc.* **433**, 3017–3047 (2013).
56. Smith, R. J., Lucey, J. R. & Conroy, C. The SINFONI nearby elliptical lens locator survey: discovery of two new low-redshift strong lenses and implications for the initial mass function in giant early-type galaxies. *Mon. Not. R. Astron. Soc.* **449**, 3441–3457 (2015).
57. Martín-Navarro, I., Barbera, F. L., Vazdekis, A., Falcón-Barroso, J. & Ferreras, I. Radial variations in the stellar initial mass function of early-type galaxies. *Mon. Not. R. Astron. Soc.* **447**, 1033–1048 (2015).
58. Milosavljević, M. & Merritt, D. Formation of galactic nuclei. *Astrophys. J.* **563**, 34–62 (2001).
59. Trujillo, I., Erwin, P., Asensio Ramos, A. & Graham, A. W. Evidence for a new elliptical-galaxy paradigm: Sérsic and core galaxies. *Astron. J.* **127**, 1917–1942 (2004).
60. Lauer, T. R. *et al.* The centers of early-type galaxies with HST. I. An observational survey. *Astron. J.* **110**, 2622–2654 (1995).
61. Carollo, C. M., Franx, M., Illingworth, G. D. & Forbes, D. A. Ellipticals with kinematically distinct cores: (V–I) color images with WFC2. *Astrophys. J.* **481**, 710–734 (1997).
62. Gerhard, O. E. Line-of-sight velocity profiles in spherical galaxies: breaking the degeneracy between anisotropy and mass. *Mon. Not. R. Astron. Soc.* **265**, 213–230 (1993).
63. van der Marel, R. P. & Franx, M. A new method for the identification of non-Gaussian line profiles in elliptical galaxies. *Astrophys. J.* **407**, 525–539 (1993).
64. Quinlan, G. D. & Hernquist, L. The dynamical evolution of massive black hole binaries—II. Self-consistent N-body integrations. *New Astron.* **2**, 533–554 (1997).



Extended Data Figure 1 | The M_{BH}/σ correlation. The black-hole masses, M_{BH} , and host-galaxy velocity dispersions, σ_e , of the 21 core galaxies (red dots, with 1σ error bars) shown in Figs 3 and 4. The dashed, dotted and solid lines show recent fits (from refs 10, 11 and 12, respectively) to

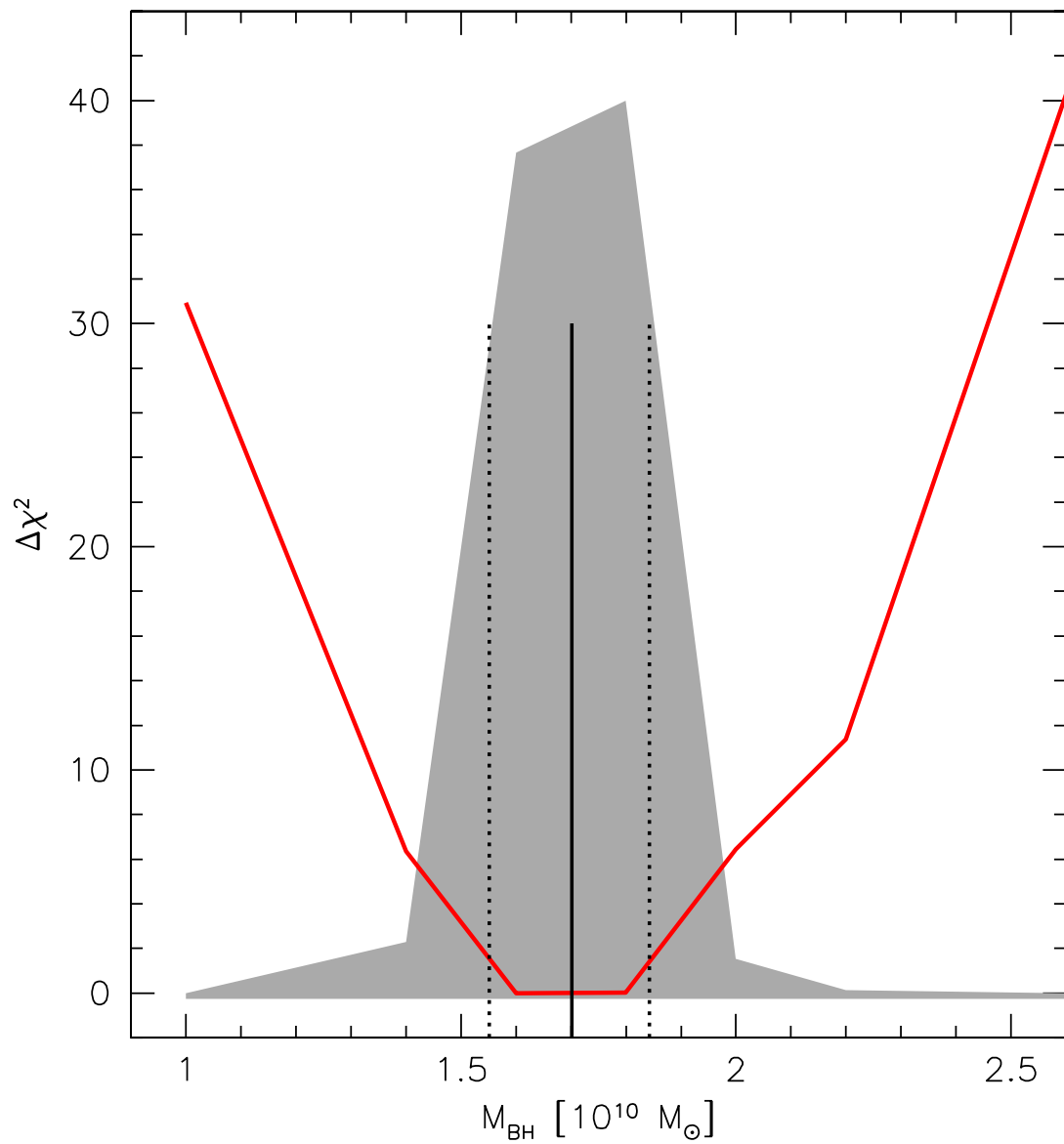
the $M_{\text{BH}}-\sigma$ correlation for all early-type galaxies (including both cored and coreless galaxies) and classical bulges with dynamically measured M_{BH} . The black hole in NGC 1600 is ten times more massive than would be expected given the galaxy's velocity dispersion ($\sigma_e = 293 \text{ km s}^{-1}$).



Extended Data Figure 2 | Surface brightness profile of NGC 1600.

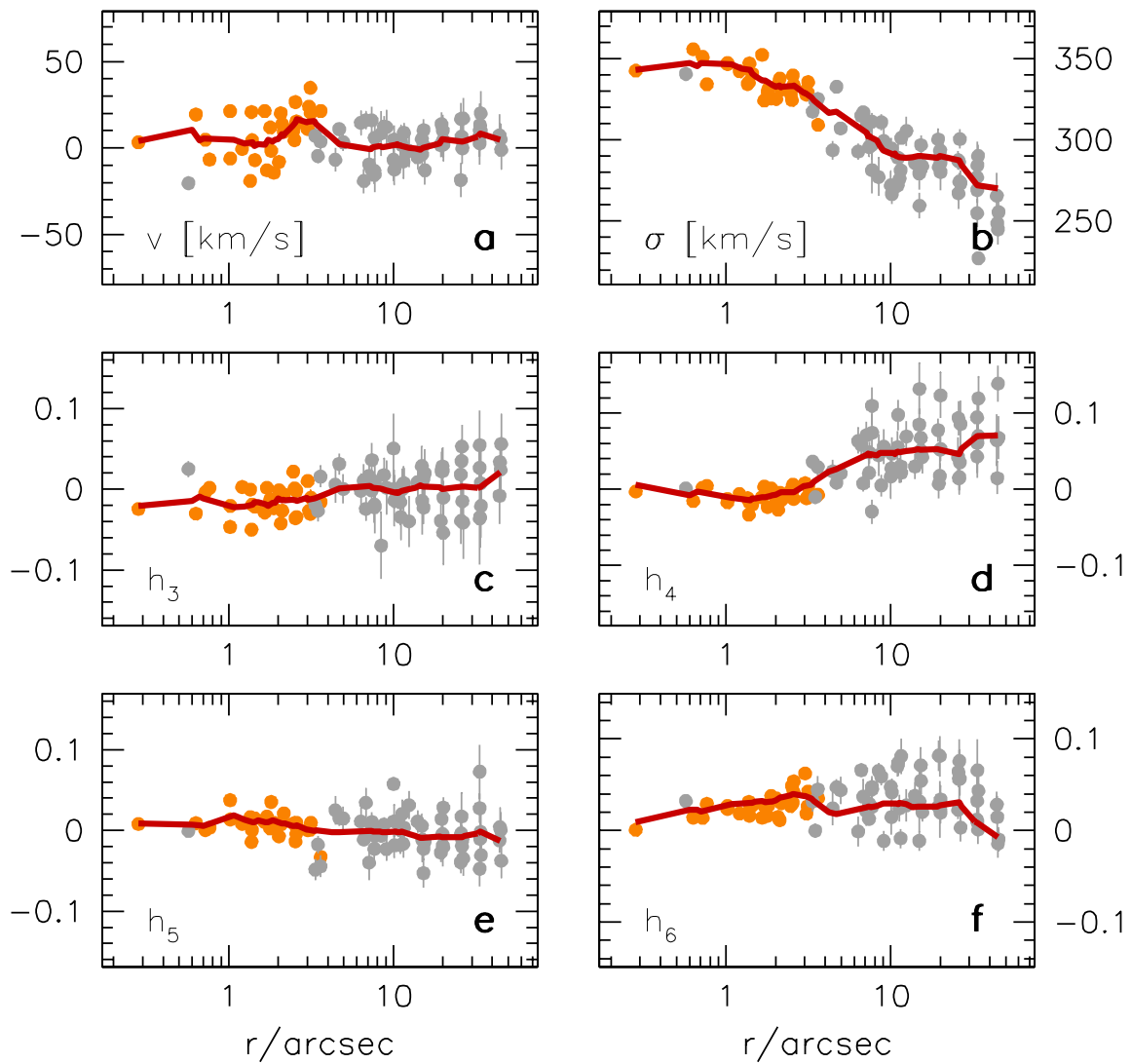
a, The circularized surface brightness distribution of NGC 1600 (filled circles) and the best-fit core-Sérsic model (red line; the best-fit parameters of the core-Sérsic function are quoted). The blue dotted line indicates the inward extrapolation of the outer Sérsic component. From the integrated

difference between the blue and the red curves, we derive a 'light deficit' of $L_{\text{def}} = 9.47 \times 10^9 L_{\odot}$. **b**, The difference between the data points in panel **a** and the core-Sérsic fit. Surface brightnesses are given in mag arcsec⁻² in the R-band.



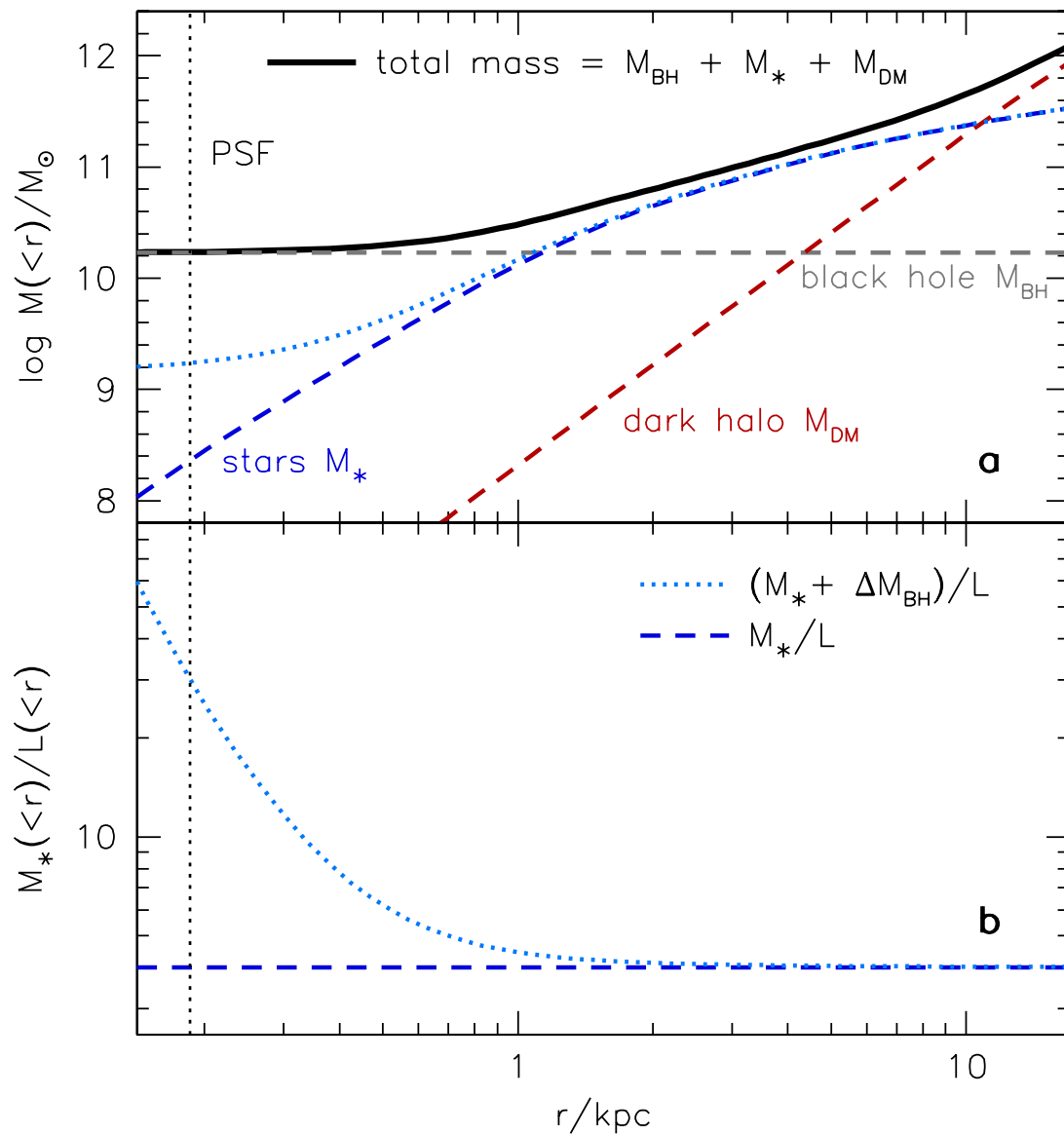
Extended Data Figure 3 | Best-fit M_{BH} values for NGC 1600 and confidence intervals. The relative likelihood of different M_{BH} values, marginalized over M_*/L , r_{DM} and v_{DM} (shaded area; the likelihood is arbitrarily scaled). The best-fit values and confidence intervals are derived

from the cumulative likelihood distribution⁴³ and indicated by the vertical lines. The red line shows $\Delta\chi^2(M_{\text{BH}}) = \chi^2(M_{\text{BH}}) - \chi_0^2$, where $\chi^2(M_{\text{BH}})$ is the minimum of all models with the same M_{BH} , but different M_*/L , r_{DM} and v_{DM} ; χ_0^2 is the minimum of $\chi^2(M_{\text{BH}})$ over M_{BH} .



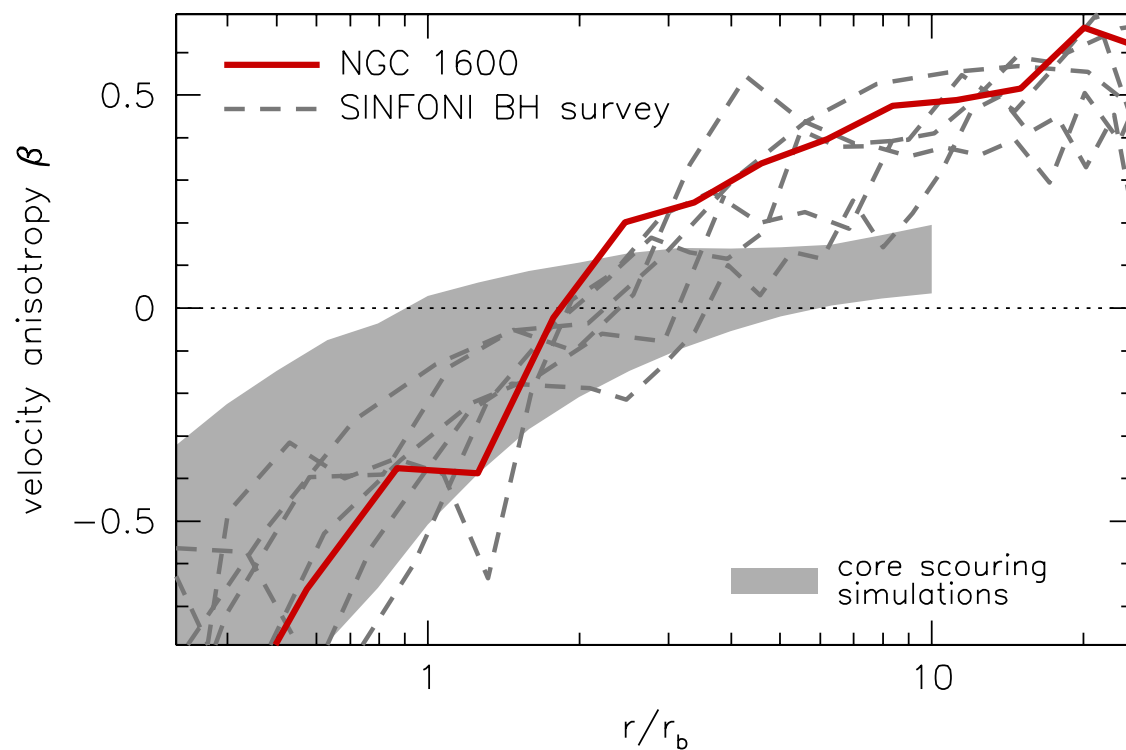
Extended Data Figure 4 | Stellar velocity data and best-fit dynamical model. These data are shown for NGC 1600 (filled grey and orange circles, with 1σ error bars), together with the best-fit model (smoothed over 0.05 dex in log radius; solid red curves). Observed LOSVDs of galaxies are approximately Gaussian and are commonly parameterized by a Gauss–Hermite series expansion^{62,63}. The mean stellar velocity v (in **a**)

and velocity dispersion σ (in **b**) correspond to the centre and the width, respectively, of the best Gaussian approximation. Higher-order Hermite coefficients h_n (in **c–f**) quantify deviations from a pure Gaussian LOSVD. Most data points at $r < 4$ arcsec came from our GMOS IFS observations (orange dots). Data at larger radii came from our Mitchell IFS observations (grey dots).



Extended Data Figure 5 | The enclosed mass of NGC 1600. **a**, The enclosed stellar mass (M_{*} , blue), dark-halo mass (M_{DM} , red), black-hole mass (M_{BH} , grey) and combined total mass (black) obtained in our model from the smallest resolved radius (point-spread-function, PSF, size) out to 20 kpc (Mitchell IFU size). **b**, An illustration of the excessive M_{*}/L gradient (dotted pale blue curve) that would be required for a hypothetical

population of unresolved central dwarf stars to explain 10% of NGC 1600's measured M_{BH} . The stellar mass-to-light ratio would have to increase by about a factor of ten (dotted pale blue curve) over our best-fit constant value (dashed blue curve). Observations of other galaxies suggest that extreme populations of dwarf stars can increase M_{*}/L by a factor of up to three.



Extended Data Figure 6 | The anisotropy of stellar orbits in core galaxies. In NGC 1600 (red line) and similar galaxies with cores^{12,18} (grey lines), the stellar velocity distribution is anisotropic. The anisotropy parameter, $\beta = 1 - \sigma_t^2/\sigma_r^2$, is positive when most of the stars move along radially stretched orbits, and negative when the stellar orbits are

predominantly tangential. Inside the diffuse, low-surface-brightness core region ($r \leq r_b$), tangential motions dominate. The shaded area indicates the range of anisotropies found in numerical N -body simulations of the core scouring mechanism^{58,64}.

Detection of a Cooper–pair density wave in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$

M. H. Hamidian^{1*}, S. D. Edkins^{2,3*}, Sang Hyun Joo^{4,5*}, A. Kostin², H. Eisaki⁶, S. Uchida^{6,7}, M. J. Lawler^{2,8}, E.-A. Kim², A. P. Mackenzie^{3,9}, K. Fujita¹⁰, Jinho Lee^{4,5} & J. C. Séamus Davis^{2,3,10,11}

The quantum condensate of Cooper pairs forming a superconductor was originally conceived as being translationally invariant. In theory, however, pairs can exist with finite momentum Q , thus generating a state with a spatially modulated Cooper-pair density^{1,2}. Such a state has been created in ultracold ⁶Li gas³ but never observed directly in any superconductor. It is now widely hypothesized that the pseudogap phase⁴ of the copper oxide superconductors contains such a ‘pair density wave’ state^{5–21}. Here we report the use of nanometre-resolution scanned Josephson tunnelling microscopy^{22–24} to image Cooper pair tunnelling from a d -wave superconducting microscope tip to the condensate of the superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. We demonstrate condensate visualization capabilities directly by using the Cooper-pair density variations surrounding zinc impurity atoms²⁵ and at the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ crystal supermodulation²⁶. Then, by using Fourier analysis of scanned Josephson tunnelling images, we discover the direct signature of a Cooper-pair density modulation at wavevectors $Q_p \approx (0.25, 0)2\pi/a_0$ and $(0, 0.25)2\pi/a_0$ in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. The amplitude of these modulations is about five per cent of the background condensate density and their form

factor exhibits primarily s or s' symmetry. This phenomenology is consistent with Ginzburg–Landau theory^{5,13,14} when a charge density wave^{5,27} with d -symmetry form factor^{28–30} and wavevector $Q_C = Q_p$ coexists with a d -symmetry superconductor; it is also predicted by several contemporary microscopic theories for the pseudogap phase^{18–21}.

When hole doping suppresses the cuprate antiferromagnetic insulator state the pseudogap regime emerges⁴. Both the high-temperature superconducting state and a charge density wave (CDW) state²⁷ exist therein. However, a ‘pair density wave’ (PDW) state, in which the Cooper-pair density modulates spatially at wavevector Q , is also widely hypothesized^{5–21} to exist in the pseudogap regime. There are compelling theoretical motivations for a cuprate PDW state. First, microscopic theories for local electronic structure of the hole-doped antiferromagnet predict a cuprate PDW state linked to combined modulations of hole density and antiferromagnetic spin density^{6–10}. Second, c -axis superconductivity does not appear in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ until temperatures far below the point at which the CuO_2 planes superconduct, a situation that could be explained by the existence of an orthogonal PDW state in each CuO_2 plane^{5,11,13}. Third, the existence of a PDW

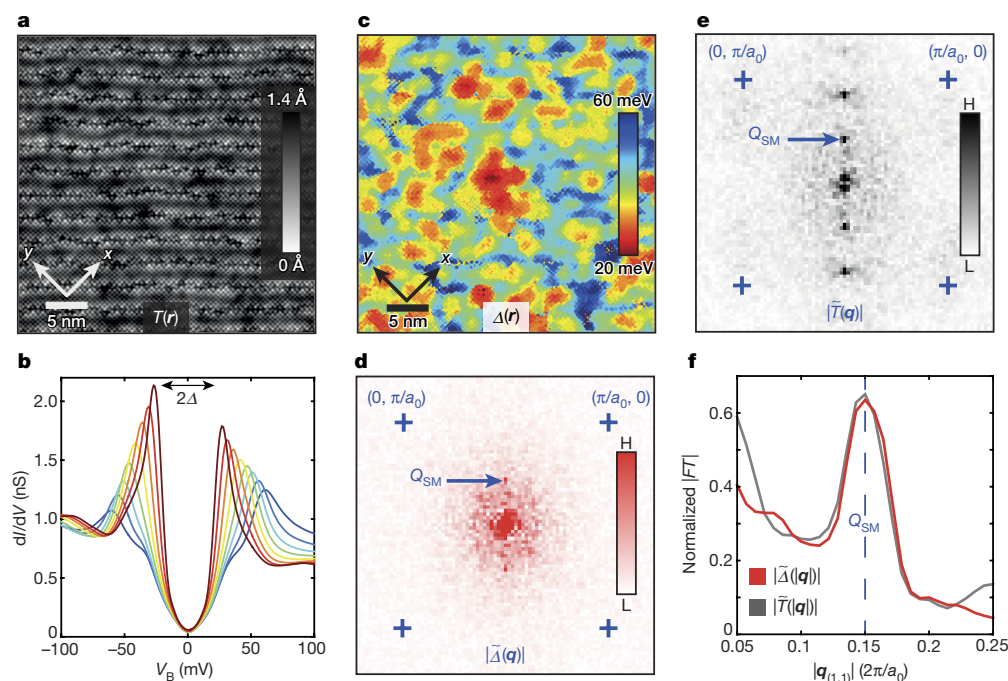


Figure 1 | Spatial variations and modulations in cuprate energy gaps. **a**, Typical $35\text{ nm} \times 35\text{ nm}$ topographic image $T(r)$ of the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ surface (crystal ‘supermodulation’ runs vertically). **b**, Typical differential conductance spectra of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. The maximum energy gap Δ is half the distance between peaks. V_B is the bias voltage. **c**, Spatial arrangement of $\Delta(r)$ (the gapmap) for $p \approx 17\%$ $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ in the same field of view as **a**. **d**, The magnitude of the Fourier transform of **c**, $|\tilde{\Delta}(q)|$ (blue crosses are at $q = (\pi/a_0, 0)$; $(0, \pi/a_0)$), showing a single peak due to the supermodulation²⁶ (blue arrow). H, high; L, low. **e**, The magnitude of the Fourier transform of topograph $|\tilde{T}(q)|$ (blue crosses are at $q = (\pi/a_0, 0)$; $(0, \pi/a_0)$). The supermodulation is a quasi-periodic modulation along the (1,1) direction with wavevector Q_{SM} indicated by the blue arrow. **f**, Simultaneously measured $\tilde{\Delta}(q)$ and $\tilde{T}(q)$ from **d** and **e** along the (1,1) direction. Their primary peaks coincide. FT, Fourier transform; normalization is to DC component.

¹Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. ²Laboratory of Atomic and Solid State Physics, Department of Physics, Cornell University, Ithaca, New York 14853, USA. ³School of Physics and Astronomy, University of St Andrews, Fife KY16 9SS, UK. ⁴Institute of Applied Physics, Department of Physics and Astronomy, Seoul National University, Seoul 151-747, South Korea. ⁵Center for Correlated Electron Systems, Institute of Basic Science, Seoul 151-742, South Korea. ⁶Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8568, Japan. ⁷Department of Physics, University of Tokyo, Bunkyo, Tokyo 113-0011, Japan. ⁸Department of Physics, Binghamton University, Binghamton, New York 13902-6000, USA. ⁹Max Planck Institute for Chemical Physics of Solids, D-01187 Dresden, Germany. ¹⁰Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA. ¹¹Kavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, New York 14853, USA.

*These authors contributed equally to this work.

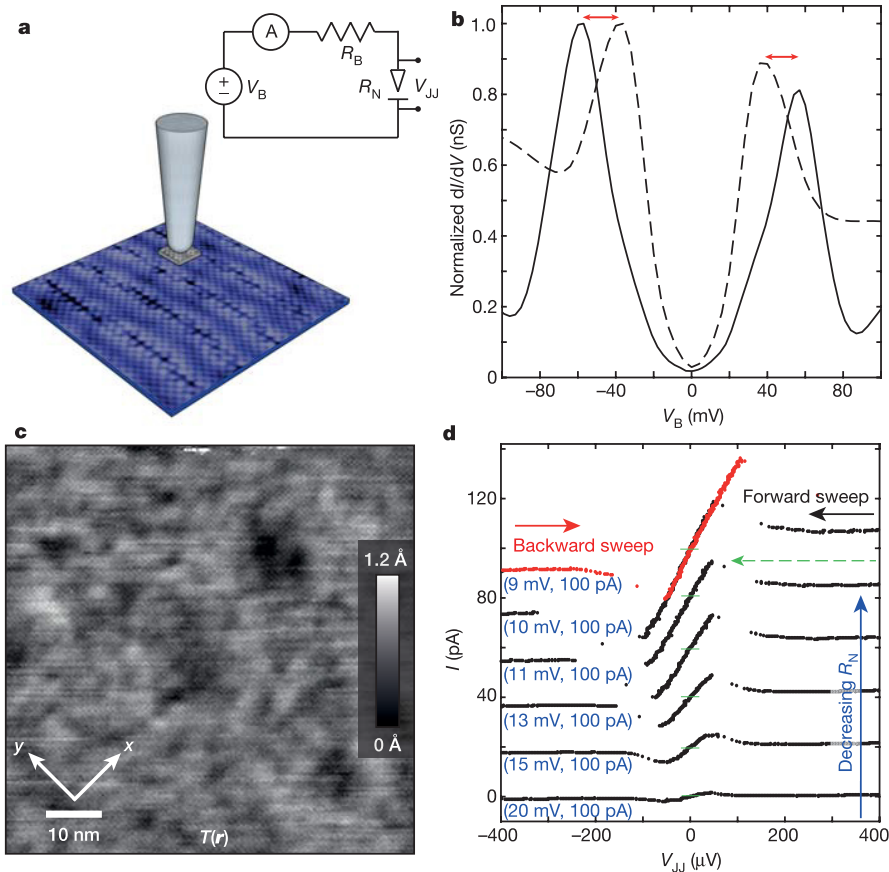


Figure 2 | *d*-wave HTS tip fabrication for SJTM. **a**, Schematic STM tip with nanometre-sized $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ flake adhering. Inset, combined spectroscopic-imaging STM and Josephson circuitry used ($R_B = 10\text{ M}\Omega$). **b**, Conversion of single-particle NIS (dashed) to single-particle SIS (solid) spectra when $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nano-flake adheres to tungsten tip. The tip gap $\Delta_T \approx 25\text{ meV}$ is the difference between the NIS and SIS peaks (red double-headed arrows). **c**, Topographic image $T(r)$ ($76\text{ nm} \times 76\text{ nm}$) using single-particle SIS tunnelling with the same tip as in **a** and **b**. Spatial

resolution is $\sim 1\text{ nm}$. Absence of moiré patterns in its Fourier transform (see Methods section ‘Characterizing $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nano-flake STM tips’) indicates that the nanoflake’s crystal axes (superconducting order parameter) are aligned with the bulk crystal axes (superconducting order parameter). **d**, Measured evolution of the SJTM $I(V)$ with diminishing tip-sample distance (diminishing R_N) at $T = 45\text{ mK}$ (junction formation conditions given in blue text). The maximum current I_c for a typical $I(V)$ is indicated by a dashed green arrow.

state could explain some of the unusual characteristics observed in the single-particle excitations^{15–17}. Finally, contemporary theories for the cuprate pseudogap phase hypothesize a composite state in which the CDW is directly linked to a PDW of the same wavevector^{18–21}.

The definitive signature^{1,2,5} of a PDW would be a periodic modulation of the Cooper-pair condensate density with wavevector \mathbf{Q} . In principle, this could be visualized using two techniques. First, scanning tunnelling microscopy (STM) using conventional single-electron tunnelling may detect modulations at \mathbf{Q} in the locally defined energy gap:

$$\Delta(\mathbf{r}) = \Delta_0 + \Delta_p \cos(\mathbf{Q} \cdot \mathbf{r}) \quad (1)$$

Here Δ_0 is the homogeneous superconductor energy gap and $\Delta_p \cos(\mathbf{Q} \cdot \mathbf{r})$ is that of the PDW (refs 1 and 2). A second possibility is scanned Josephson tunnelling microscopy (SJTM) of Cooper pairs^{22–24} to detect the modulations of Cooper-pair density directly. Again, in the presence of the homogeneous condensate, this should result in a modulating Josephson critical current I_J :

$$I_J(\mathbf{r}) = I_J^0 + I_J^p \cos(\mathbf{Q} \cdot \mathbf{r}) \quad (2)$$

where the first term represents Cooper-pair tunnelling to a homogeneous condensate and the second to the PDW. Imaging the spatial arrangements of the superfluid density $\rho_s(\mathbf{r})$ would then be possible, since $I_J^2(\mathbf{r}) \propto \rho_s(\mathbf{r})$ (see Methods section ‘Modelling phase diffusion Josephson dynamics for SJTM’).

The phenomena in equation (1) have not been detected in a cuprate (or any other) superconductor. One challenge is the extreme variation in the differential tunnelling conductance spectra (Fig. 1b) typically detected at the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ surface $T(r)$ (Fig. 1a). Visualizing the maximum energy gap (double-headed black arrow in Fig. 1b) yields a heterogeneous ‘gapmap’ $\Delta(\mathbf{r})$ (Fig. 1c), whose Fourier transform $\tilde{\Delta}(\mathbf{q})$ (Fig. 1d) exhibits only one sharp finite- \mathbf{q} peak (blue arrow). This $\Delta(\mathbf{r})$ modulation occurs at the wavevector \mathbf{Q}_{SM} of a quasi-periodic distortion of the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ crystal unit cell along the CuO_2 (1,1) direction²⁶ that is referred to as the ‘supermodulation’ (Fig. 1e). Because no other periodic modulations of the maximum energy gap Δ (or any other energy gap) have been detected in cuprate gapmap studies, searches for a cuprate PDW via equation (1) yield a null result.

The use of superconducting STM tips to measure the spatial variation of Cooper-pair tunnelling (SJTM) therefore remains the most promising approach with which to search for a PDW state (equation (2)). For two superconducting electrodes with temperature dependent energy gaps $\Delta(T)$ and phases ϕ_1 and ϕ_2 , the Josephson current is given by $I(\phi) = I_J \sin(\phi)$ where $\phi = \phi_2 - \phi_1$. The Josephson critical current I_J is given by $I_J R_N = \frac{\pi \Delta(T)}{2e} \tanh(\Delta(T)/2kT)$ where R_N is the normal-state junction resistance, k is the Boltzmann constant, and $2e$ is the Cooper-pair charge. Time independence of ϕ would require the Josephson energy $E_J = \frac{\phi_0 I_J}{2\pi}$, where ϕ_0 is the magnetic flux quantum, to exceed the thermal fluctuation energy kT . For a nanometre-sized Josephson junction with sub-nanoampere I_J , this regime remains out of reach at

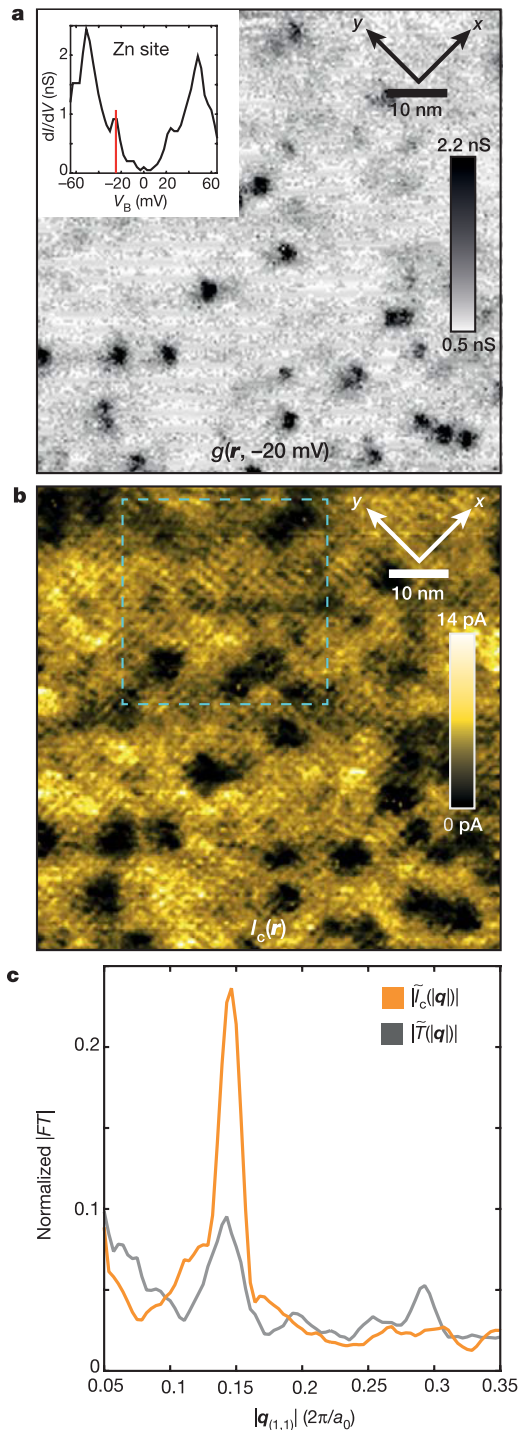


Figure 3 | Cooper-pair condensate visualization using SJTM. **a**, The convolution of $\Omega = -1.5$ meV Zn impurity resonance with the d -wave spectrum of the tip shifts the resonance to $|E| \approx |\Omega| + |\Delta_T|$. The measured SIS $g(\mathbf{r}, E = -20$ meV) $76 \text{ nm} \times 76 \text{ nm}$ image near this energy (vertical red line) identifies the site of each Zn. **b**, A $76 \text{ nm} \times 76 \text{ nm}$ $I_c(\mathbf{r})$ image measured at 45 mK in the same field of view as **a**. A deep minimum ($\sim 95\%$ suppression) occurs in $I_c(\mathbf{r})$ surrounding each Zn site where the Cooper-pair condensate is suppressed²⁵. **c**, Simultaneously measured magnitude of $\tilde{I}_c(\mathbf{q})$ and $\tilde{T}(\mathbf{q})$ along the CuO_2 (1,1) direction. The primary peaks in $\tilde{I}_c(\mathbf{q})$ and from the supermodulation coincide, demonstrating the ability to visualize Cooper-pair density modulations.

millikelvin temperatures. However, imaging of phase-diffusion-dominated Josephson tunnelling is possible by using superconducting-tip STM to measure an $I(V)$ whose maximum current is $I_c \propto I_J^2$

(refs 22–24 and Methods section ‘Modelling phase diffusion Josephson dynamics for SJTM’).

To search for a cuprate PDW near $\mathbf{Q} \approx (0.25, 0)2\pi/a_0$ and $(0, 0.25)2\pi/a_0$ where $a_0 \approx 3.8 \text{ \AA}$ also requires visualizing $I_c(\mathbf{r}) \propto I_J^2(\mathbf{r})$ with nanometre resolution. Moreover, high Δ , low R_N and millikelvin operating temperatures would all be highly advantageous. Motivated in this way, we use a dilution-refrigerator-based STM operated below 50 mK, and achieve a high tip-energy gap Δ_T by picking up a nanometre-sized flake of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ on the end of our tungsten tip (see Fig. 2a and Methods section ‘Characterizing $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nano-flake STM tips’). This immediately converts our measured normal–insulator–superconductor (NIS) tunnelling spectrum (dashed line in Fig. 2b) to a superconductor–insulator–superconductor (SIS) spectrum (solid line in Fig. 2b) exhibiting an energy separation of about $2(\Delta_S + \Delta_T)$ between conductance peaks (Methods section ‘Characterizing $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nano-flake STM tips’). Figure 2b demonstrates that our d -wave high-temperature superconducting (HTS) tip has $\Delta_T \approx 25$ meV (double-headed red arrows), and its spatial resolution is evaluated to be 1 nm using a topographic image of the BiO surface of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ (Fig. 2c) that was measured using this tip in the single-particle tunnelling regime (Methods section ‘Characterizing $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nanoflake STM tips’). Next, we measure the $I(V)$ characteristic of this tip as a function of decreasing tip distance from the bulk crystal surface, and thus of decreasing R_N . At $T = 45$ mK and fixed \mathbf{r} , as the superconducting tip is moved forward in ~ 10 -picometre steps the evolution of $I(V)$ characteristics is as shown in Fig. 2d. As the Josephson I_J increases with diminishing distance, the maximum observable current $I_c \propto I_J^2$ increases as expected (see Methods section ‘Modelling phase diffusion Josephson dynamics for SJTM’ and ‘Characterizing $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nanoflake STM tips’).

To test the Cooper-pair condensate visualization capabilities of this $I_c(\mathbf{r})$ imaging technique directly, we use two approaches. First, when Zn impurity atoms are substituted at the Cu sites, muon spin rotation studies show that the Cooper-pair condensate is completely suppressed in a radius of a few nanometres around each Zn atom, in a ‘Swiss cheese’ configuration²⁵. Figure 3a shows a $76 \text{ nm} \times 76 \text{ nm}$ image of SIS single-particle tunnelling conductance $g(\mathbf{r}, -20$ meV) using the same tip. One can locate each Zn impurity atom by its scattering resonance peak, which is shifted by convolution with the superconducting tip spectrum from its resonance energy $\Omega = -1.5$ meV to $|E| \approx |\Omega| + |\Delta_T| \approx 25$ meV (vertical red line in the inset to Fig. 3a). Thus, the Zn impurity atoms are at the dark spots in Fig. 3a. In Fig. 3b we show $I_c(\mathbf{r})$ measured at ~ 50 mK in the same field of view as Fig. 3a. This reveals how $I_c(\mathbf{r})$, and thus the Cooper-pair condensate, is suppressed within a radius $R \approx 3$ nm of each Zn atom site, consistent with muon spin rotation studies²⁵. A second test is possible because of the crystal ‘supermodulation’ and the associated modulating superconductivity at wavevector \mathbf{Q}_{SM} (ref. 26) (Fig. 1f). In Fig. 3c we plot the magnitude of the Fourier transform $\tilde{T}(\mathbf{q})$ in grey and $\tilde{I}_c(\mathbf{q})$ in orange, both measured simultaneously along the (1,1) direction through \mathbf{Q}_{SM} . This directly demonstrates the capability to visualize the Cooper-pair density modulations that are induced by the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ crystal supermodulation. Overall, these tests show that nanometre-resolution Cooper-pair condensate visualization is achievable using this d -wave high-gap tip, combined with our millikelvin-operating-temperature SJTM approach.

Next, we apply this technique to search for a cuprate PDW in the same $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ samples with $T_c = 88$ K and hole density $p = 17\%$ and at $T < 50$ mK (see Methods section ‘Defining, measuring and imaging $I_c(\mathbf{r})$ ’). The two states already reported to coexist at this p value are the high-temperature superconductor and a CDW with $\mathbf{Q} = (0.22 \pm 0.02, 0)2\pi/a_0$ and $(0, 0.22 \pm 0.02)2\pi/a_0$ (ref. 27). Figure 4a shows our measured $I_c(\mathbf{r})$ in the $35 \text{ nm} \times 35 \text{ nm}$ field of view (blue dashed box in Fig. 3b). Clear $I_c(\mathbf{r})$ modulations are immediately observable and, given that $I_c(\mathbf{r}) \propto (I_J^0)^2 + 2I_J^0 I_J^P \cos(\mathbf{Q} \cdot \mathbf{r}) + (I_J^P)^2 \cos^2(\mathbf{Q} \cdot \mathbf{r})$

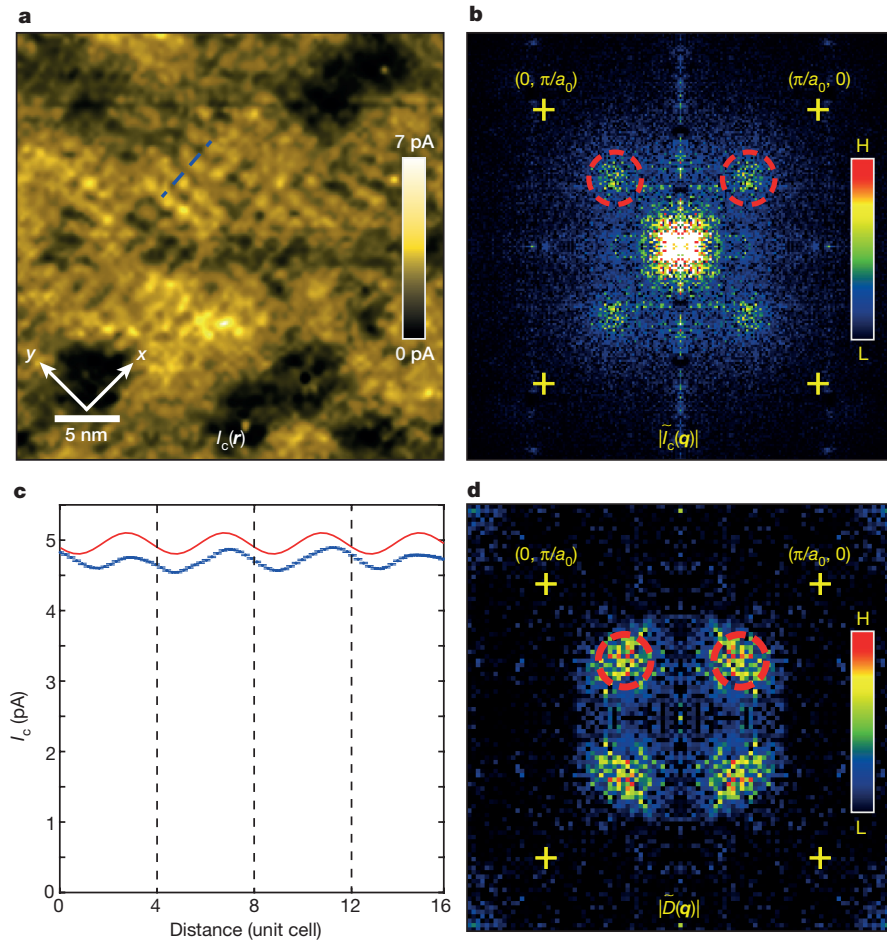


Figure 4 | Visualizing the Cooper-pair density wave in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. **a**, A $35 \text{ nm} \times 35 \text{ nm}$ field of view $I_c(\mathbf{r})$ image (dashed blue box in Fig. 3b) showing the $I_c(\mathbf{r})$ modulations parallel to the CuO_2 x and y axes (supermodulation-induced $I_c(\mathbf{r})$ modulations are removed). **b**, $|\tilde{I}_c(\mathbf{q})|$, the Fourier transform of $I_c(\mathbf{r})$ in **a** (yellow crosses at $\mathbf{q} = (\pi/a_0, 0)$ and $(0, \pi/a_0)$). Maxima from $I_c(\mathbf{r})$ modulations (dashed red circles) occur at $\mathbf{Q}_p = (0.25, 0)2\pi/a_0$ and $(0, 0.25)2\pi/a_0$, and no significant modulations occur in $\tilde{R}_N(\mathbf{Q}_p)$ (see Methods section ‘Visualizing supermodulation-

from equation (2), they provide the magnitude and wavevector of the component $I_j^p(\mathbf{r})$ modulations. The magnitude of the Fourier transform of $I_c(\mathbf{r})$, $|\tilde{I}_c(\mathbf{q})|$, is shown in Fig. 4b, revealing that modulation in $I_c(\mathbf{r})$ occurs at the wavevectors $\mathbf{Q}_p = (0.25 \pm 0.02, 0)2\pi/a_0$ and $(0, 0.25 \pm 0.02)2\pi/a_0$ (dashed red circles in Fig. 4b), while nothing is observable in the second harmonic $\cos^2(\mathbf{Q} \cdot \mathbf{r})$. This situation seems to occur because the $I_j^p(\mathbf{r})$ modulations are superposed on a much stronger spatially constant critical current I_j^0 (equation (2)). These data provide strong *a priori* evidence for the existence of a PDW coexisting with a robust Cooper-pair condensate in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$. In Fig. 4c we show (as blue dots) the measured value of $I_c(\mathbf{r})$ along the blue line in Fig. 4a, while we show the amplitude and wavelength of the global $I_c(\mathbf{r})$ modulations (determined from the magnitude and central \mathbf{Q} value of the peaks in Fig. 4b) as a red line. From these data, and in general from the magnitude of the peaks at \mathbf{Q}_x and \mathbf{Q}_y in Fig. 4b, we conclude that the modulation amplitude in Cooper-pair density is $\sim 5\%$ of its average value. Therefore the cuprate PDW at $p = 17\%$ is subdominant to the d -wave superconducting state below 50 mK .

Finally, in Fig. 4d we study $\tilde{D}(\mathbf{q}) = \tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$, the oxygen-sublattice-phase-resolved image of d -symmetry form factor density modulations^{28–30}, from a $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ sample with the same hole density (see Methods section ‘Imaging d -symmetry CDW $\tilde{D}(\mathbf{q})$ and avoiding the CDW setup effect in $I_c(\mathbf{r})$ ’). The conventional locations of

induced PDW and $R_N(\mathbf{r})$). **c**, Measured $I_c(\mathbf{r})$ along the dashed blue line in **a** (blue dotted line); statistical error bars are the variance of I_c transverse to the dashed line. The red line shows the global amplitude and \mathbf{Q}_p of the $I_c(\mathbf{r})$ modulations (from the circled maxima in **b**). **d**, Magnitude of $\tilde{D}(\mathbf{q}) = \tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$ (yellow crosses at $\mathbf{q} = (\pi/a_0, 0)$ and $(0, \pi/a_0)$) revealing the d -symmetry form factor density wave modulations (dashed red circles) at $\mathbf{Q}_C = (0.22, 0)2\pi/a_0$; $(0, 0.22)2\pi/a_0$ (same hole density as in **b**).

d -symmetry form factor CDW peaks in $\tilde{D}(\mathbf{q})$ (dashed red circles) occur at $\mathbf{Q}_C = (0.22 \pm 0.04, 0)2\pi/a_0$ and $(0, 0.22 \pm 0.04)2\pi/a_0$. We note that in the unprocessed data the modulations in electronic structure and topography occur based upon the sum of both oxygen sublattices ($\tilde{O}_x(\mathbf{q}) + \tilde{O}_y(\mathbf{q})$) and at wavevector $\mathbf{Q} = \mathbf{Q}_{\text{Bragg}} \pm \mathbf{Q}_C$ (ref. 28). No response of the tip to these modulations could produce a spurious $I_c(\mathbf{r})$ modulation at the PDW wavevector $\mathbf{Q}_p = (0.25 \pm 0.02, 0)2\pi/a_0$ and $(0, 0.25 \pm 0.02)2\pi/a_0$ (see Methods section ‘Imaging d -symmetry CDW $\tilde{D}(\mathbf{q})$ and avoiding the CDW setup effect in $I_c(\mathbf{r})$ ’). Obviously, Fig. 4b and Fig. 4d are not identical, with the PDW exhibiting much narrower peaks and thus more spatial coherence. Nevertheless, this PDW wavevector (Fig. 4b, c) is not inconsistent (within joint error bars) with a conventionally defined wavevector of the cuprate CDW state at the same hole density²⁷. However, the observed PDW also exhibits a primarily s/s' -symmetry form factor because Fig. 4b is based on the conventional sum over sublattices, while the CDW state exhibits a primarily d -symmetry form factor, as detected in the difference of sublattice-phase-resolved images²⁸ $\tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$ (Fig. 4d).

The visualization of a Cooper-pair density wave, a long-term challenge in physics^{1,2}, has now been achieved using SJTM^{22–24}. The subdominant PDW observed in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ at approximately the same wavevector as a CDW ($\mathbf{Q}_p \approx \mathbf{Q}_C$) and with s/s' form factor (Fig. 4), is consistent with Ginzburg–Landau theory^{5,11,13,14} in the case

(among others) where a d -symmetry superconductor coexists with a d -symmetry form factor CDW (see Methods section 'Analysing PDW within Ginzburg–Landau theory'). Microscopic theories for a cuprate PDW^{9,13,16} requiring $Q_p = Q_C/2$ cannot be fully tested here because of the observed disorder in $\tilde{I}_c(q)$ at low q (Fig. 4b). However, microscopic models for the pseudogap phase involving the interplay of d -symmetry Cooper pairing and a d -symmetry form factor CDW do yield a PDW with $Q_p \approx Q_C$ (refs 18–21) as in Fig. 4, although typically with a d -symmetry form factor. Moreover, the apparently lattice-commensurate value of Q_p , which would arise naturally in strong-coupling theories for density modulations in a doped Mott insulator, motivates further study. Finally, our HTS-tip millikelvin-SJTM approach to nanometre-resolution Cooper-pair condensate imaging also reveals a direct route to visualization of FFLO^{1,2} or PDW^{6–21} states in other cuprates, iron pnictides and unconventional superconductors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 November 2015; accepted 8 February 2016.

Published online 13 April 2016.

- Fulde, P. & Ferrell, R. A. Superconductivity in a strong spin-exchange field. *Phys. Rev.* **135**, A550–A563 (1964).
- Larkin, A. I. & Ovchinnikov, Y. N. Neodnorodnoe sostoyanie sverkhprovodnikov. *Zh. Eksp. Teor. Fis.* **47**, 1136–1146 (1964); Inhomogeneous state of superconductors. *Sov. Phys. JETP* **20**, 762–769 (1965).
- Liao, Y. *et al.* Spin-imbalance in a one-dimensional Fermi gas. *Nature* **467**, 567–569 (2010).
- Norman, M. R. & Pépin, C. The electronic nature of high temperature cuprate superconductors. *Rep. Prog. Phys.* **66**, 1547–1610 (2003).
- Fradkin, E., Kivelson, S. A. & Tranquada, J. M. Theory of intertwined orders in high temperature superconductors. *Rev. Mod. Phys.* **87**, 457–482 (2015).
- Himeda, A., Kato, T. & Ogata, M. Stripe states with spatially oscillating d -wave superconductivity in the two-dimensional t - t' - J model. *Phys. Rev. Lett.* **88**, 117001 (2002).
- Raczkowski, M., Capello, M., Poilblanc, D., Frésard, R. & Oleś, A. M. Unidirectional d -wave superconducting domains in the two-dimensional t - J model. *Phys. Rev. B* **76**, 140505 (2007).
- Yang, K.-Y., Chen, W. Q., Rice, T. M., Sigrist, M. & Zhang, F.-C. Nature of stripes in the generalized t - J model applied to the cuprate superconductors. *New J. Phys.* **11**, 055053 (2009).
- Loder, F., Graser, S., Kampf, A. P. & Kopp, T. Mean-field pairing theory for the charge-stripe phase of high-temperature cuprate superconductors. *Phys. Rev. Lett.* **107**, 187001 (2011).
- Corboz, P., Rice, T. M. & Troyer, M. Competing states in the t - J model: Uniform d -wave state versus stripe state. *Phys. Rev. Lett.* **113**, 046402 (2014).
- Berg, E. *et al.* Dynamical layer decoupling in a stripe-ordered high- T_c superconductor. *Phys. Rev. Lett.* **99**, 127003 (2007).
- Seo, K., Chen, H.-D. & Hu, J. Complementary pair-density-wave and d -wave-checkerboard orderings in high-temperature superconductors. *Phys. Rev. B* **78**, 094510 (2008).
- Berg, E., Fradkin, E. & Kivelson, S. A. Charge- $4e$ superconductivity from pair-density-wave order in certain high-temperature superconductors. *Nature Phys.* **5**, 830–833 (2009).
- Agterberg, D. F. & Tsunetsugu, H. Dislocations and vortices in pair-density-wave superconductors. *Nature Phys.* **4**, 639–642 (2008).
- Zelli, M., Kallin, C. & Berlinsky, A. J. Quantum oscillations in a π -striped superconductor. *Phys. Rev. B* **86**, 104507 (2012).
- Lee, P. A. Amperean pairing and the pseudogap phase of cuprate superconductors. *Phys. Rev. X* **4**, 031017 (2014).
- Chen, H.-D., Vafeek, O., Yazdani, A. & Zhang, S.-C. Pair density wave in the pseudogap state of high temperature superconductors. *Phys. Rev. Lett.* **93**, 187002 (2004).
- Pépin, C., de Carvalho, V. S., Kloss, T. & Montiel, X. Pseudogap, charge order, and pairing density wave at the hot spots in cuprate superconductors. *Phys. Rev. B* **90**, 195207 (2014).
- Freire, H., de Carvalho, V. S. & Pépin, C. Renormalization group analysis of the pair-density-wave and charge order within the fermionic hot-spot model for cuprate superconductors. *Phys. Rev. B* **92**, 045132 (2015).
- Wang, Y., Agterberg, D. F. & Chubukov, A. Interplay between pair- and charge-density-wave orders in underdoped cuprates. *Phys. Rev. B* **91**, 115103 (2015).
- Wang, Y., Agterberg, D. F. & Chubukov, A. Coexistence of charge-density-wave and pair-density-wave orders in underdoped cuprates. *Phys. Rev. Lett.* **114**, 197001 (2015).
- Naaman, O., Teizer, W. & Dynes, R. C. Fluctuation dominated Josephson tunneling with a scanning tunneling microscope. *Phys. Rev. Lett.* **87**, 097004 (2001).
- Rodrigo, J. G., Suderow, H. & Vieira, S. On the use of STM superconducting tips at very low temperatures. *Eur. Phys. J. B* **40**, 483–488 (2004).
- Proslir, Th. *et al.* Probing the superconducting condensate on a nanometer scale. *Europhys. Lett.* **73**, 962–968 (2006).
- Nachumi, B. *et al.* Muon spin relaxation studies of Zn-substitution effects in high- T_c cuprate superconductors. *Phys. Rev. Lett.* **77**, 5421 (1996).
- Slezak, J. A. *et al.* Imaging the impact on cuprate superconductivity of varying the interatomic distances within individual crystal unit cells. *Proc. Natl Acad. Sci. USA* **105**, 3203–3208 (2008).
- Comin, R. & Damascelli, A. Resonant X-ray scattering studies of charge order in cuprates. *Ann. Rev. Cond. Matter Phys.* **7**, 369–405 (2016).
- Fujita, K. *et al.* Direct phase-sensitive identification of a d -form factor density wave in underdoped cuprates. *Proc. Natl Acad. Sci. USA* **111**, E3026–E3032 (2014).
- Comin, R. *et al.* Symmetry of charge order in cuprates. *Nature Mater.* **14**, 796–801 (2015).
- Forgan, E. M. *et al.* The microscopic structure of charge density waves in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6.54}$ revealed by X-ray diffraction. *Nature Comm.* **6**, 10064 (2015).

Acknowledgements We acknowledge and thank D. Agterberg, A. V. Balatsky, D. Chowdhury, A. Chubukov, E. Fradkin, R. Hulet, S. A. Kivelson, P. A. Lee, M. Norman, J. W. Orenstein, C. Pepin, S. Sachdev, J. Tranquada and Y. Wang for discussions and advice. The development and operation of HTS SJTM technology and M.H.H. and A.K. were funded by the Moore Foundation's EPiQS Initiative through grant number GBMF4544. S.D.E. acknowledges studentship funding from the EPSRC under grant number EP/G03673X/1. J.C.S.D. and A.P.M. acknowledge research support from the EPSRC through the grant programme 'Topological Protection and Non-Equilibrium States in Correlated Electron Systems'. S.U. and H.E. acknowledge support from a Grant-in-Aid for Scientific Research from the Ministry of Science and Education (Japan). S.H.J. and J.L. acknowledge support from the Institute for Basic Science, Korea under grant number IBS-R009-D1. J.C.S.D. and K.F. acknowledge salary support from the US Department of Energy, Office of Basic Energy Sciences, under contract number DEAC02-98CH10886. E.-A.K. acknowledges support from the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering under award DE-SC0010313.

Author Contributions M.H.H., S.D.E., A.K., and J.L. developed the SJTM techniques and carried out the experiments. K.F., H.E. and S.U. synthesized and characterized the samples. M.H.H., S.D.E., A.K., S.H.J. and K.F. developed and carried out analyses. E.-A.K. and M.J.L. provided theoretical guidance. A.P.M., J.L. and J.C.S.D. supervised the project and wrote the paper with key contributions from M.H.H., S.D.E. and K.F. The manuscript reflects the contributions and ideas of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.S.D. (jcseamusdavis@gmail.com) or J.L. (jinhlee@snu.ac.kr).

METHODS

Modelling phase diffusion Josephson dynamics for SJTM. The Josephson effect³¹ is first described by

$$I = I_J \sin(\phi) \quad (3)$$

the relationship of the Cooper-pair current I through the junction to the phase difference ϕ between the two electrodes, with I_J the Josephson critical current. Second,

$$\frac{d\phi}{dt} = \frac{2e}{\hbar} V \quad (4)$$

relates the time evolution of that phase difference to the voltage across the junction. If the Josephson coupling energy $E_J = \frac{\hbar}{2e} I_J \gg kT$ so that effects of thermal fluctuations on ϕ can be ignored, there exists a steady-state solution to equations (3) and (4) where $V = 0$ for $I < I_J$. At 50 mK, the condition $E_J = kT$ corresponds to a Josephson critical current $I_J \approx 2$ nA.

Because of our operating junction resistance $R_N \approx 10$ M Ω , the maximum supercurrent used in the SJTM experiments described here was approximately 20 pA. Thus, our experimental junctions all lie in a regime where ϕ exhibits steady-state diffusion through the Josephson potential owing to thermal fluctuations^{32–38}. This results in a voltage dropped across the junction while the Cooper-pair current is flowing. We analyse the Josephson $I(V)$ characteristics between our d -wave HTS tip and the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ (BSCCO) surface using this classic phase-diffusion model for small junctions^{32,33} within which the $I(V)$ characteristic of the junction is given by

$$I(V_J) = \frac{1}{2} I_J^2 Z \frac{V_J}{V_J^2 + V_c^2} \quad (5)$$

V_J is the voltage across the junction and $V_c = \frac{2eZkT^*}{\hbar}$ with Z the impedance in series with the voltage source at the high frequencies relevant to repeated re-trapping of the diffusing phase. Importantly, $I(V_J)$ is maximal at

$$I_c = \frac{\hbar}{8ekT^*} I_J^2 \quad (6)$$

where T^* is an effective temperature parameterizing how dissipative the phase evolution is. Spatially resolved measurements of $I_c(\mathbf{r})$ thus allow one to measure spatial modulations in $I_J^2(\mathbf{r})$ (refs 22–24).

Extended Data Fig. 1a shows a diagram of the hybrid spectroscopic-imaging STM/SJTM measurement circuit used. A load resistor $R_B = 10$ M Ω is voltage-biased in series with the Josephson junction formed between the STM tip and the sample. Extended Data Fig. 1b shows the zero-frequency solutions to an electronic circuit model for this setup. The red and blue lines show the trajectories that the forwards and backwards sweeps take through the $I(V)$ plane. The nonlinearity of the circuit introduces discontinuities in the measured $I(V)$; a discontinuity occurs at the point where $I = I_c$ and thus makes this current experimentally identifiable by the sharp $I(V)$ feature associated with it. Extended Data Fig. 1c shows the same trajectories but as a function of the voltage across the junction V_J . This circuit's nonlinearity also generates hysteresis in the $I(V)$ characteristic even if the dynamics of the Josephson junction itself are over-damped at all frequencies. This hysteresis is indeed observed throughout our SJTM studies of BSCCO (Fig. 2d), and systematic errors due to the hysteresis are avoided during our $I_c(\mathbf{r})$ imaging experiments by sweeping the applied voltage always in the same direction once the Josephson junction has been formed at each location \mathbf{r} .

Characterizing $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ nanoflake STM tips. To detect the gap modulations (equation (1)) requires single electron tunnelling from a non-superconducting electrode (NIS), while to observe the Josephson current modulations (equation (2)) requires Cooper-pair tunnelling from a superconducting electrode (SIS). For our SJTM studies, the high energy gap at the STM tip Δ_T is achieved by picking up a nanometre-sized flake of BSCCO onto the conventional tungsten microscope tip. Extended Data Fig. 1d shows the measured dI/dV spectrum of the tungsten tip before that process as a dashed line; it is precisely as expected for hole density $p \approx 17\%$ BSCCO samples. Simulation of the expected dI/dV for tunnelling between a BSCCO nanoflake tip and sample (solid line) is in good agreement with the typical SIS spectrum observed.

In Extended Data Fig. 2 we compare the performance of two completely different BSCCO nanoflake tips. Extended Data Fig. 2a and b shows their measured dI/dV spectra; they are obviously very similar. Extended Data Fig. 2c and d demonstrates the ~ 1 -nm spatial resolution achieved in topographic images using these two tips. Extended Data Fig. 2e and f shows the Fourier transform of these two topographic images, demonstrating that the supermodulation is easily detected in both cases and that no other unusual characteristics, such as additional q -space peaks or moiré patterns, are observed.

To check the relative orientation of the BSCCO nanoflake crystal axes to the axes of the sample, we evaluate simulations to see what might be expected of such a tip. Extended Data Fig. 3a and b shows a measured NIS topograph $T(\mathbf{r})$ of BSCCO taken with a conventional tungsten tip, and its Fourier transform. To simulate what would be observed if the crystal axes of a BSCCO nanoflake are oriented (or not) to those of the sample crystal axes we extract a typical ~ 2 -nm-diameter patch from this image and then convolute the patch image with $T(\mathbf{r})$ at each location \mathbf{r} . This results in the simulated topograph that is expected when using a BSCCO nanoflake tip with (or without) axes aligned to the crystal axes. If the axes of the patch are aligned (Extended Data Fig. 3c) to the crystal, the result is as in Extended Data Fig. 3e (see the Fourier transform in the inset). If the axes are misaligned (Extended Data Fig. 3d) with the crystal the result is as in Extended Data Fig. 3f (see the Fourier transform in the inset). As the vivid moiré effects in the Fourier transform of the latter are not observed in the actual experiments (Extended Data Fig. 2e and f), we conclude that the axes of the nanoflake are aligned with those of the BSCCO bulk crystal.

Specifications for HTS-tip, millikelvin SJTM for condensate visualization. The technical challenge of large field-of-view nanometre-resolution SJTM imaging is to measure, in a reasonable time (days), an array of $I_c(\mathbf{r})$ values with $\sim 1\%$ precision using typically 256×256 pixels. Four basic elements are required: (1) an ultra-low-vibration dilution-refrigerator-based STM that is engineered for sub-picometre stability in taking spectroscopic data while out of feedback, and typically at $T = 50$ mK; (2) a high-temperature superconducting tip creation scheme which involves removing a nanometre-sized flake from the BSCCO surface and adhering it to the tip; (3) modification to the electronic circuitry to form a hybrid spectroscopic-imaging STM/SJTM operating condition (Fig. 2a) that allows conventional operation in the SIS configuration for topographic imaging and motion control, and bias voltage sweeps in the microvolt range when in Josephson mode; and (4) control of the capability to reliably and repeatedly carry out spectroscopic maps $I(V, \mathbf{r})$ using a junction resistance of ~ 10 M Ω at every \mathbf{r} and with V spanning tens of microvolts.

Defining, measuring and imaging $I_c(\mathbf{r})$. Extended Data Fig. 4a shows the histogram of I_c values measured in a typical field of view during these studies. By averaging over the individual measured $I(V)$ from all the locations \mathbf{r} contained within the bins indicated by the coloured arrows, we obtain the three representative Josephson $I(V)$ measurements shown in Extended Data Fig. 4b. They span the range of $I(V)$ characteristics measured at different locations. The coloured arrows in Extended Data Fig. 4b give examples where I_c is identified in a particular curve. In general we functionally define the quantity $I_c(\mathbf{r})$ reported in the main text to be the maximum magnitude of current achieved on sweeping away from $V = 0$. By comparison with Extended Data Fig. 1b, this current is the maximum Cooper-pair current, I_c , described by the phase diffusion model (see Methods section 'Modelling phase diffusion Josephson dynamics for SJTM') that at constant temperature and Z is directly proportional to $I_J^2(\mathbf{r})$.

Extended Data Fig. 5a, b and c shows typical examples of repeated $I_c(\mathbf{r})$ measurement experiments (with slightly different operating parameters but the same BSCCO nano-flake tip). The functionality and repeatability of this technique to achieve $I_c(\mathbf{r})$ imaging with nanometre resolution is clear.

The typical evolution of measured $I(V)$ during SJTM imaging of $I_c(\mathbf{r})$ at 45 mK at a specific sequence of locations (numbered 1–9 in Extended Data Fig. 5b) is shown in Extended Data Fig. 6. The almost complete ($\sim 95\%$) suppression of I_c as this sequence passes through the site of a Zn atom is seen in panel 5 of Extended Data Fig. 6. These typical data demonstrate directly the unprocessed signal-to-noise ratio for measurement of I_c and how spatial evolution of unprocessed $I(V)$ spectra yields the high fidelity seen in the $I_c(\mathbf{r})$ images.

In Extended Data Fig. 7a and b we show the topographic images taken with the same BSCCO nanoflake tip before and after a typical $I_c(\mathbf{r})$ mapping experiment. Comparison of the topographs shows them to be virtually identical and demonstrates that no significant changes occur in the geometry of the tip during the map. The implication is that the nanoflake at the end of the tip never once touches the surface, and that all such $I_c(\mathbf{r})$ studies reported are in the true SIS tunnelling limit and are very far from making any point contact.

Finally, our model is that the BSCCO tip has maximum gap Δ_T and negligible modulating component, while that of the sample Δ_S has two components Δ_S^0 (background component) and Δ_S^p (PDW component) and $\Delta_S = \Delta_S^0 + \Delta_S^p$ with $\Delta_S^0 \gg \Delta_S^p$. The Josephson coupling then consists of a primary coupling that occurs between the uniform components, and a secondary coupling that is between the uniform component of the tip and the modulated component of the sample (as in equation (2)) with the empirical result of convolution presented in Fig. 4. To leading order $I_c(\mathbf{r})$ is then linear in the PDW component of both the sample gap and superfluid density.

Visualizing supermodulation-induced PDW and $R_N(\mathbf{r})$. In principle, SJTM allows direct measurement of modulations in the superconducting order

parameter amplitude through the product $I_J R_N$. Here, I_J is the intrinsic, $T=0$, Josephson critical current of the junction and R_N is a characteristic resistance of the single-particle tunnelling channel. In taking this product, variations that affect the matrix elements for tunnelling of Cooper pairs and the single-particle tunnelling conductance ($1/R_N$) in the same way cancel out.

We use dI/dV at a bias voltage V_B far higher than the SIS gap edge as an estimate of the ungapped, 'normal state' junction conductance G_N . Simultaneous measurement of this value and $I_c(r)$ were not possible because the former requires V_B in tens of millivolts while the latter requires V_B of a few tens of microvolts. Moreover, in the picometre range tip-sample separations used in measuring $I_c(r)$, the single-particle tunnelling current can become destructively large to our BSCCO nanoflake tip. Thus the following procedure was adopted:

(1) Measure $I_c(r)$ with setpoint current $I_s = I_{s1}$ and bias voltage $V_s = V_{s1}$, as shown in Extended Data Fig. 8a.

(2) In the same field of view measure $g_1(r, E)$, defined as dI/dV for bias voltages V_{s1} to $-V_{s1}$ with setpoint condition $I_s = I_{s1}$, $V_s = V_{s1}$.

(3) In the same field of view measure $g_2(r, E)$ defined as dI/dV with setpoint condition $I_s = I_{s2}$, $V_s = V_{s2}$ for bias voltages V_{s2} to $-V_{s2}$, where $V_{s2} \gg V_{s1}$ and $V_{s2} > \Delta_{\text{SIS}}$.

(4) Over the range $V = [-V_{s1}, V_{s1}]$ determine the coefficients $\alpha(r)$ and $\beta(r)$, which scale $g_2(r)$ onto $g_1(r)$ via $g_1(r, E) = \alpha(r)g_2(r, E) + \beta(r)$. In doing the scaling a small region of voltages near $V = 0$ where the Cooper-pair tunnelling (Josephson) signal exhibits is excluded.

(5) $R_N(r)$ for junctions used to measure $I_c(r)$ is then determined by $1/R_N(r) = G_N(r) = \alpha(r)/2(g_2(r, V_{s2}) + g_2(r, -V_{s2})) + \beta(r)$, as shown in Extended Data Fig. 8b.

Now we consider the case of the cuprate supermodulation PDW where the bulk quasi-periodic distortion of the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ crystal at wavevector \mathbf{Q}_{SM} leads to modulations of the measured $I_c(r)$ at the same wavevector superposed on the larger background.

$$\begin{aligned} \Delta(r) &\propto I_J(r)R_N(r) = (I_J^0 + I_J^p \cos(\mathbf{Q}_{\text{SM}} \cdot \mathbf{r}))(R_N^0 + R_N^p \cos(\mathbf{Q}_{\text{SM}} \cdot \mathbf{r} + \varphi)) \\ &= I_J^0 R_N^0 + R_N^0 I_J^p \cos(\mathbf{Q}_{\text{SM}} \cdot \mathbf{r}) + I_J^p R_N^p \cos(\mathbf{Q}_{\text{SM}} \cdot \mathbf{r} + \varphi) + \dots \end{aligned} \quad (7)$$

where we allow for an arbitrary phase difference φ between modulation in I_J and R_N at \mathbf{Q}_{SM} . For any value of φ , one may neglect the third term if

$$I_J^0 R_N^p \ll R_N^0 I_J^p \Leftrightarrow \frac{R_N^p}{R_N^0} \div \frac{I_J^p}{I_J^0} \ll 1 \quad (8)$$

In the experiments presented in the main text we measure the quantity

$$I_c(r) \propto (I_J^0)^2 + 2I_J^0 I_J^p \cos(\mathbf{Q}_{\text{SM}} \cdot \mathbf{r}) + \dots \quad (9)$$

as shown in Extended Data Fig. 8a. In Extended Data Fig. 8b we show $R_N(r)$ derived as above. By comparing the amplitude of $\tilde{I}_c(\mathbf{q})$ and $\tilde{R}_N(\mathbf{q})$ at $\mathbf{q} = \mathbf{Q}_{\text{SM}}$ normalized to their $\mathbf{q} = \mathbf{0}$ value one may directly determine $\frac{R_N^p}{R_N^0} \div \frac{I_J^p}{I_J^0}$ as in Extended Data Fig. 9c; we find that this ratio is far less than 1 and hence modulations in $\Delta(r)$ at \mathbf{Q}_{SM} are faithfully represented by those measured in $I_c(r)$.

Inset to Extended Data Fig. 8c is $|\tilde{R}_N(\mathbf{q})|$, showing directly that variations in the $\tilde{R}_N(\mathbf{q})$ near $\mathbf{q} \approx (0.25, 0)2\pi/a_0$ and equivalent are negligible, so that measurement of $I_c(r)$ should and does yield a faithful image of the condensate.

Imaging d -symmetry CDW $\tilde{D}(\mathbf{q})$ and avoiding the CDW setup effect in $I_c(r)$. The d -symmetry form factor density wave in underdoped cuprates (refs 39–41) has been directly detected by spectroscopic-imaging STM by analysing the differential conductance $dI/dV \equiv g(r, V)$ images in terms of $\tilde{D}(\mathbf{q}) = \tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$ which extracts the d -symmetry form factor density wave signature at the correct wavevector. At dopings near optimal, it is known from several experimental techniques that the intensity of the d -symmetry form factor density wave diminishes rapidly⁴². Therefore, to improve the visualization of the d -symmetry form factor density wave signature at such high dopings as used here, we use topographic information instead. In constant-current topographic imaging, the STM feedback system adjusts the tip-sample separation, T , as it scans over the sample surface to maintain a setpoint current, I_s , at a constant applied tip-sample bias V_s . The topographic image $T(r)$ is related to the tunnelling current

$$I(r, T, V) = f(r, T) \int_0^{eV_s} \text{LDOS}(\mathbf{r}, \epsilon) d\epsilon \quad (10)$$

and assuming that the function $f(r, T)$, which represents the effect of corrugation, work function, and tunnelling matrix elements, takes the form

$$f(r, T) = \exp(-\kappa T) A(r) \quad (11)$$

where κ is the tunnel-barrier-height factor and depends on the work functions of the sample and tip, and LDOS is the local density of electronic states. The recorded value of the relative tip-sample separation is then

$$T(r) = \frac{1}{\kappa} \ln \left[\int_0^{eV_s} \text{LDOS}(\mathbf{r}, \epsilon) d\epsilon \right] + \frac{1}{\kappa} \ln \left[\frac{I_s}{A(r)} \right] \quad (12)$$

Thus high signal-to-noise topographic images obtained by constant-current STM imaging reveal contributions from both the surface structure and variations in the LDOS(\mathbf{r}, E).

While a single cuprate topographic image will contain the signature of the d -symmetry form factor density wave, subtracting one topographic image from another taken at opposite bias polarity can further enhance the cuprate CDW signal. This is because the cuprate d -symmetry form factor density wave modulates in anti-phase⁴¹ between empty and filled states and hence a subtraction of topographic images measured near the pseudogap energy ($|V_s| \approx 100$ meV in equation (12)) formed with biases $+V_s$ and $-V_s$ should amplify the CDW contrast.

In Extended Data Fig. 9a and b we demonstrate this procedure using two high-resolution, high-signal-to-noise ratio topographs $T(r)$ taken at opposite bias polarities (100 mV and -100 mV) with the same conventional tungsten tip; they are spatially registered to each other with picometre precision⁴⁰. Extended Data Fig. 9c is the difference of these topographs, from which the two oxygen-sublattice-sampled images $O_x(r)$; $O_y(r)$ are then extracted, and $\tilde{D}(\mathbf{q}) = \tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$ calculated. An image of $\tilde{D}(\mathbf{q})$ reveals a primarily d -symmetry form factor density wave because of the existence of four maxima at the CDW wavevector $\mathbf{Q}_C = (\pm 0.22, 0)2\pi/a_0$ and $(0, \pm 0.22)2\pi/a_0$ as in Fig. 4d (refs 39 and 41).

One important implication of the fact that the topographic/CDW modulations have d -symmetry form factor is for avoidance of the setup effect when measuring $I_c(r)$. Equation (12) shows that $T(r)$ is affected logarithmically by the term $\int_0^{eV_s} \text{LDOS}(\mathbf{r}, \epsilon) d\epsilon$. This means the tip-sample distance will be modulated at the same \mathbf{Q} as modulations in LDOS(\mathbf{r}, ϵ). One concern might be that, because the CDW modulations produce $T(r)$ (that is, the tip-sample distance) modulations at this \mathbf{Q} , this will generate a spurious $I_c(r)$ modulation at the identical \mathbf{Q} . This would be the infamous 'setup effect', but now affecting SJTM.

However, because the CDW in cuprates has a predominant d -symmetry form factor, any conventional image $T(r)$ and its Fourier transform $\tilde{T}(\mathbf{q})$ that are formed by adding the three sublattice images exhibit the actual modulations at wavevector $\mathbf{Q} = \mathbf{Q}_{\text{Bragg}} \pm \mathbf{Q}_C = (1 \pm 0.22, 0)2\pi/a_0$ and $(0, 1 \pm 0.22)2\pi/a_0$. These only become detectable at the correct CDW wavevector \mathbf{Q}_C when one uses a measure of d -symmetry form factor: $\tilde{D}(\mathbf{q}) = \tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$. Therefore CDW modulations in the unprocessed topograph $T(r)$; $\tilde{T}(\mathbf{q})$ occur at $\mathbf{Q} \approx (0.78, 0)2\pi/a_0$ and $(0, 0.78)2\pi/a_0$ (dashed red circles in the Extended Data Fig. 10a). Because they are at a completely different wavevector, it is impossible for such modulations to produce, through a setup effect, spurious $I_c(r)$ modulations at $\mathbf{Q} \approx (0.25, 0)2\pi/a_0$ and $(0, 0.25)2\pi/a_0$. Therefore the PDW wavevector \mathbf{Q}_P observed directly by SJTM, and shown in the Extended Data Fig. 10b and in Fig. 4b, is not generated spuriously by a systematic setup effect due to the coexisting CDW.

Analysing PDW within Ginzburg–Landau theory. PDW modulations may be induced by the coupling between a translationally invariant superconducting order parameter and a CDW. We focus on the case of a tetragonal system with $\mathbf{q} = \mathbf{0}$ superconducting order of B_{1g} ($d_{x^2-y^2}$) symmetry, a CDW order parameter with a purely d -symmetry form factor⁴³ and an induced secondary PDW order parameter with s/s' symmetry form factor. The pairing amplitude is expanded in terms of superconducting order parameters via

$$\begin{aligned} \phi(\mathbf{r}, \mathbf{r}') &\equiv \psi_d(\mathbf{r})\psi_{-d}(\mathbf{r}') = D(\mathbf{r} - \mathbf{r}')\Delta_0(\mathbf{R}) \\ &+ S'(\mathbf{r} - \mathbf{r}')\{\Delta_Q(\mathbf{R})e^{i\mathbf{Q}\cdot\mathbf{R}} + \Delta_{-Q}(\mathbf{R})e^{-i\mathbf{Q}\cdot\mathbf{R}} + \Delta_{\bar{Q}}(\mathbf{R})e^{i\bar{\mathbf{Q}}\cdot\mathbf{R}} + \Delta_{-\bar{Q}}(\mathbf{R})e^{-i\bar{\mathbf{Q}}\cdot\mathbf{R}}\} \end{aligned} \quad (13)$$

where $\mathbf{R} = (\mathbf{r} - \mathbf{r}')/2$, $D(\mathbf{r} - \mathbf{r}')$ and $S'(\mathbf{r} - \mathbf{r}')$ are the d - and s' -form factors that either do or do not change sign under 90° rotations. This expansion then defines five complex order parameters that enter the free energy: Δ_0 , Δ_Q , Δ_{-Q} , $\Delta_{\bar{Q}}$ and $\Delta_{-\bar{Q}}$. Δ_Q and Δ_{-Q} are required to be independent to implement the U(1) spatial phase degree of freedom because $\phi(\mathbf{r}, \mathbf{r}')$ itself is complex. \mathbf{Q} and $\bar{\mathbf{Q}}$ are along symmetry-equivalent orthogonal directions. Under a 90° rotation the B_{1g} superconducting and s/s' form factor PDW order parameters transform as $\Delta_0 \rightarrow -\Delta_0$ and $\Delta_Q \rightarrow \Delta_{\bar{Q}}$, $\Delta_{\bar{Q}} \rightarrow \Delta_Q$, $\Delta_{-Q} \rightarrow \Delta_{-\bar{Q}}$, $\Delta_{-\bar{Q}} \rightarrow \Delta_{-Q}$.

In a similar way, the CDW order parameter is

$$\begin{aligned} \phi_c(\mathbf{r}, \mathbf{r}') &\equiv \langle \psi_\sigma^\dagger(\mathbf{r})\psi_\sigma(\mathbf{r}') + \psi_{\sigma'}^\dagger(\mathbf{r}')\psi_{\sigma'}(\mathbf{r}) \rangle \\ &= D(\mathbf{r} - \mathbf{r}')\{\psi_Q(\mathbf{R})e^{i\mathbf{Q}\cdot\mathbf{R}} + \psi_{-Q}(\mathbf{R})e^{-i\mathbf{Q}\cdot\mathbf{R}} + \psi_{\bar{Q}}(\mathbf{R})e^{i\bar{\mathbf{Q}}\cdot\mathbf{R}} + \psi_{-\bar{Q}}(\mathbf{R})e^{-i\bar{\mathbf{Q}}\cdot\mathbf{R}}\} \\ &\quad \{\psi_Q(\mathbf{R})e^{i\mathbf{Q}\cdot\mathbf{R}} + \psi_{-Q}(\mathbf{R})e^{-i\mathbf{Q}\cdot\mathbf{R}} + \psi_{\bar{Q}}(\mathbf{R})e^{i\bar{\mathbf{Q}}\cdot\mathbf{R}} + \psi_{-\bar{Q}}(\mathbf{R})e^{-i\bar{\mathbf{Q}}\cdot\mathbf{R}}\} \end{aligned} \quad (14)$$

This will contribute two independent complex order parameters to the free energy expansion. Since the CDW order parameter must be real $\psi_{-\mathbf{Q}} = \psi_{\mathbf{Q}}^*$. For purely d -symmetry form factor CDW, the order parameter will have the following property under a 90° rotation: $\psi_{\mathbf{Q}} \rightarrow -\psi_{\bar{\mathbf{Q}}}$, $\psi_{\bar{\mathbf{Q}}} \rightarrow -\psi_{-\mathbf{Q}}$, $\psi_{-\mathbf{Q}} \rightarrow -\psi_{-\bar{\mathbf{Q}}}$, $\psi_{-\bar{\mathbf{Q}}} \rightarrow -\psi_{\mathbf{Q}}$. Now consider the symmetry-allowed free energy terms that will couple Δ_0 , the $\Delta_{\mathbf{Q}}$ values and the $\psi_{\mathbf{Q}}$ values. The leading term is simply:

$$\gamma_1[\Delta_0^*\{\psi_{\mathbf{Q}}\Delta_{-\mathbf{Q}} + \psi_{-\mathbf{Q}}\Delta_{\mathbf{Q}} + \psi_{\bar{\mathbf{Q}}}\Delta_{-\bar{\mathbf{Q}}} + \psi_{-\bar{\mathbf{Q}}}\Delta_{\bar{\mathbf{Q}}}\} + c \cdot c] \quad (15)$$

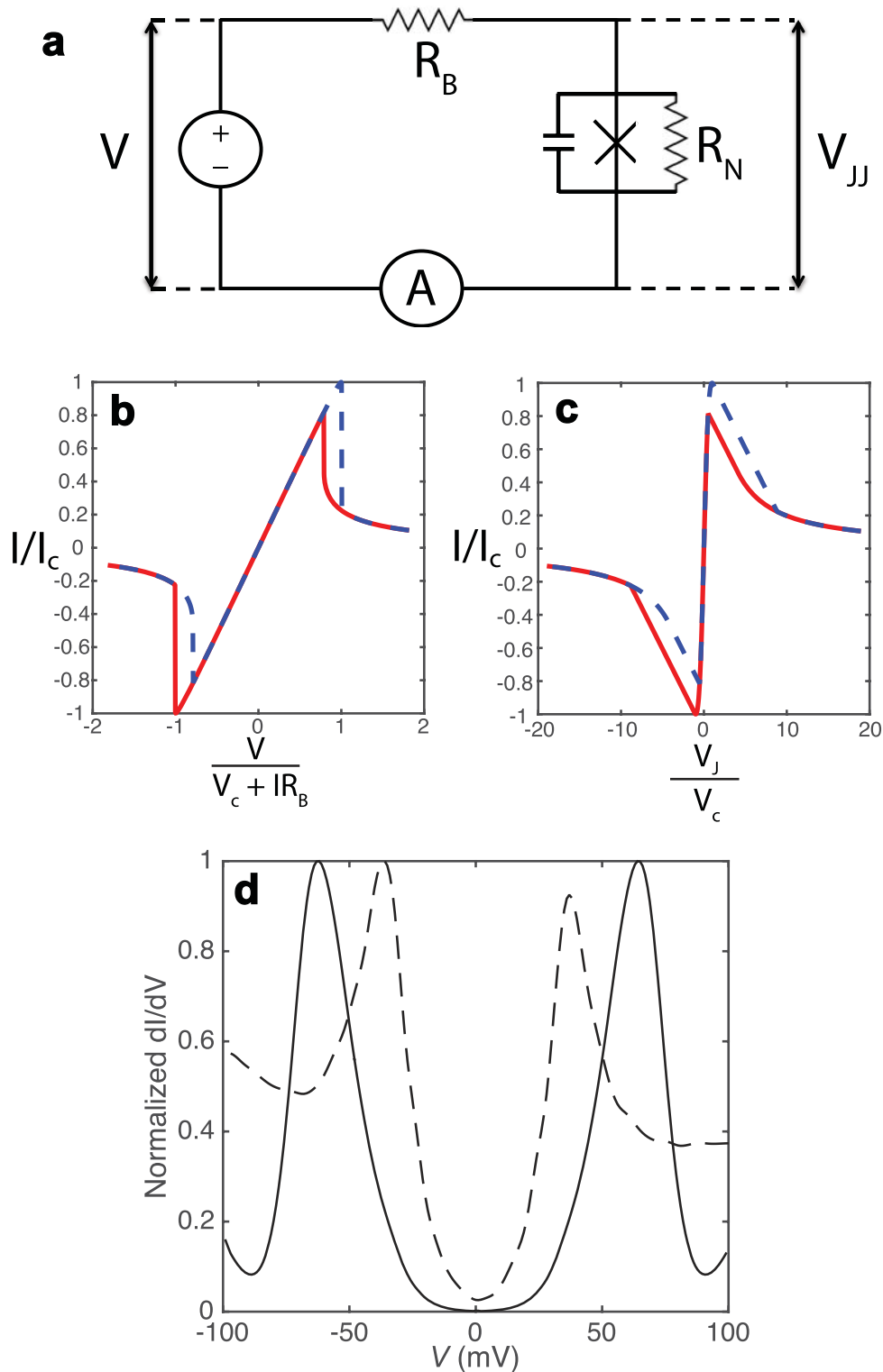
Because the PDW order parameters appear to the power of unity in this term, the free energy can always be lowered by making $\Delta_{\mathbf{Q}}$ non-zero, if both Δ_0^* and $\psi_{\bar{\mathbf{Q}}}$ are non-zero. Thus the co-existence of $d_{x^2-y^2}$ superconductivity and a d -symmetry form factor CDW at wave-vector \mathbf{Q} will induce a PDW at wavevector \mathbf{Q} with an s/s' symmetry form factor.

We note that such a coupling is not unique. There exists a similar symmetry-allowed term in the free energy admitting a d -symmetry form factor PDW where the $\Delta_{\mathbf{Q}}$ values now transform under 90° degree rotation, like the $\psi_{\mathbf{Q}}$ values. The associated free energy term will have a relative negative sign between the first two terms and the second two terms of equation (15). As the experimentally detected PDW has s/s' form factor we do not discuss this further here.

Code availability. The data files for the results presented here and the codes used in their analysis are available at <http://dx.doi.org/10.17630/83957a84-2186-4c14-8e55-f961a19ec9a9>.

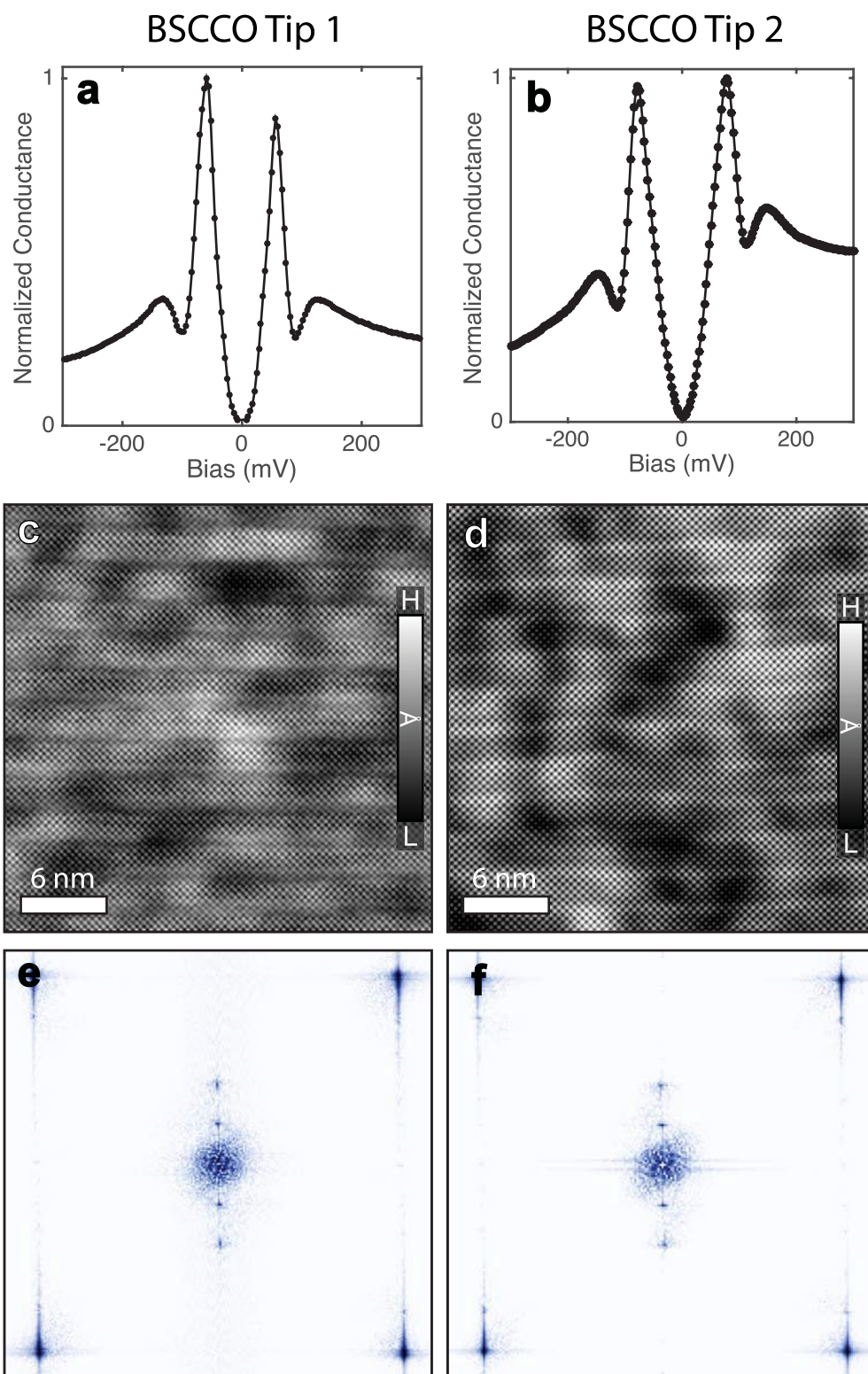
31. Josephson, B. D. Possible new effects in superconductive tunneling. *Phys. Lett.* **1**, 251–253 (1962).

32. Ambegaokar, V. & Halperin, B. I. Voltage due to thermal noise in the dc Josephson effect. *Phys. Rev. Lett.* **22**, 1364–1366 (1969); erratum **23**, 274 (1969).
33. Ivanchenko, Y. M. & Zil'berman, L. A. The Josephson effect in small tunnel contacts. *Sov. Phys. JETP* **28**, 1272–1276 (1969).
34. Devoret, M. H. *et al.* Effect of the electromagnetic environment on the Coulomb blockade in ultrasmall tunnel junctions. *Phys. Rev. Lett.* **64**, 1824–1827 (1990).
35. Falci, G., Bujanja, V. & Schön, G. Quantum tunnelling in small-capacitance Josephson junctions in a general electromagnetic environment. *Europhys. Lett.* **16**, 109–114 (1991).
36. Martinis, J. M. & Kautz, R. L. Classical phase diffusion in small hysteretic Josephson junctions. *Phys. Rev. Lett.* **63**, 1507–1510 (1989).
37. Iansiti, M., Tinkham, M., Johnson, A. T., Smith, W. F. & Lobb, C. J. Charging effects and quantum properties of small superconducting tunnel junctions. *Phys. Rev. B* **39**, 6465 (1989).
38. Šmakov, J., Martin, I. & Balatsky, A. V. Josephson scanning tunneling microscopy. *Phys. Rev. B* **64**, 212506 (2001).
39. Fujita, K. *et al.* Spectroscopic imaging scanning tunneling microscopy studies of electronic structure in the superconducting and pseudogap phases of cuprate high- T_c superconductors. *J. Phys. Soc. Jpn.* **81**, 011005 (2012).
40. Hamidian, M. H. *et al.* Picometer registration of zinc impurity states in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ for phase determination in intra-unit-cell Fourier transform STM. *New J. Phys.* **14**, 053017 (2012).
41. Hamidian, M. H. *et al.* Atomic-scale electronic structure of the cuprate d -symmetry form factor density wave state. *Nature Phys.* **12**, 150–156 (2016).
42. Fujita, K. *et al.* Simultaneous transitions in cuprate momentum-space topology and electronic symmetry breaking. *Science* **344**, 612–616 (2014).
43. Sachdev, S. & La Placa, R. Bond order in two-dimensional metals with anti-ferromagnetic exchange interactions. *Phys. Rev. Lett.* **111**, 027202 (2013).

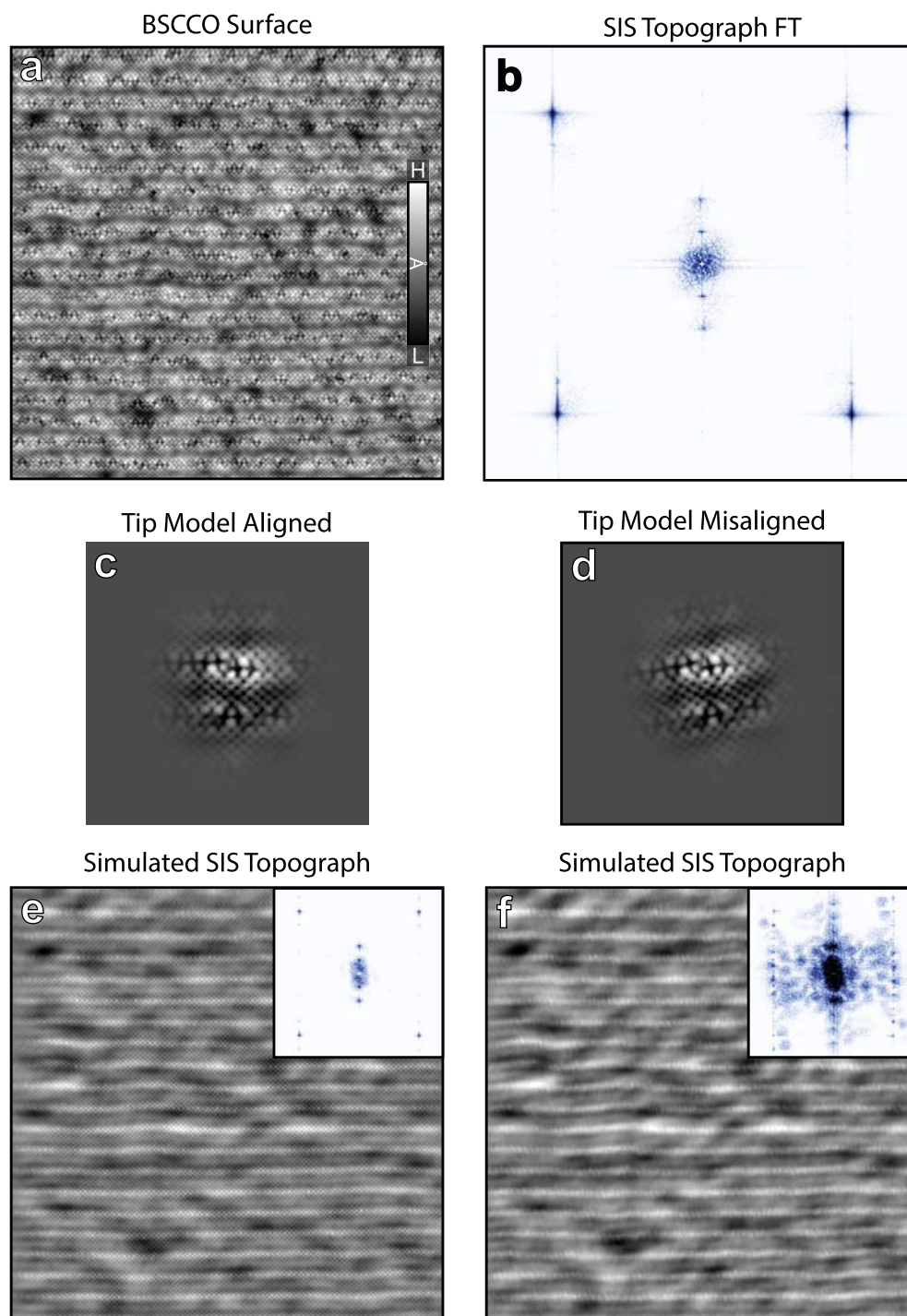


Extended Data Figure 1 | SJTM circuit model and phase diffusion dynamics. **a**, Circuit diagram of the hybrid spectroscopic-imaging STM/SJTM setup used in these experiments. The voltage source V is from the usual STM bias controller; the load resistor $R_B = 10 \text{ M}\Omega$; the single-particle tunnelling resistance of the Josephson junction formed between the tip and sample is R_N , and the voltage actually developed across the junction is V_{JJ} . **b**, The dynamics of this circuit produces two predominant effects. First there is a sudden change in the $I(V)$ characteristic measured with the external ammeter shown, when the current reaches a value $I_c = \frac{\hbar}{8ekT^*} I_J^2$ where I_J is the zero-temperature Josephson critical current and T^* is an effective temperature parameterizing the dissipative environment of the junction. The second predicted effect is strong hysteresis depending on

which direction the external voltage is swept; this is shown as the difference between the solid-red and dashed-blue lines. Both effects are seen very clearly and universally in the measured $I(V)$ throughout our studies reported here. **c**, The relationship of the current as in **b** but in terms of the voltage across the Josephson junction V_{JJ} . **d**, The dashed curve represents a typical dI/dV spectrum on as-grown BSCCO sample, measured using a normal metallic tungsten tip. The solid line is a simulation for an expected dI/dV spectrum when using a BSCCO superconducting tip; we use the standard equation for tunnelling between two superconductors, each with the density of states identical to the dashed line. The result, in very good agreement with the typical measured SIS spectrum, is shown in Fig. 2b.

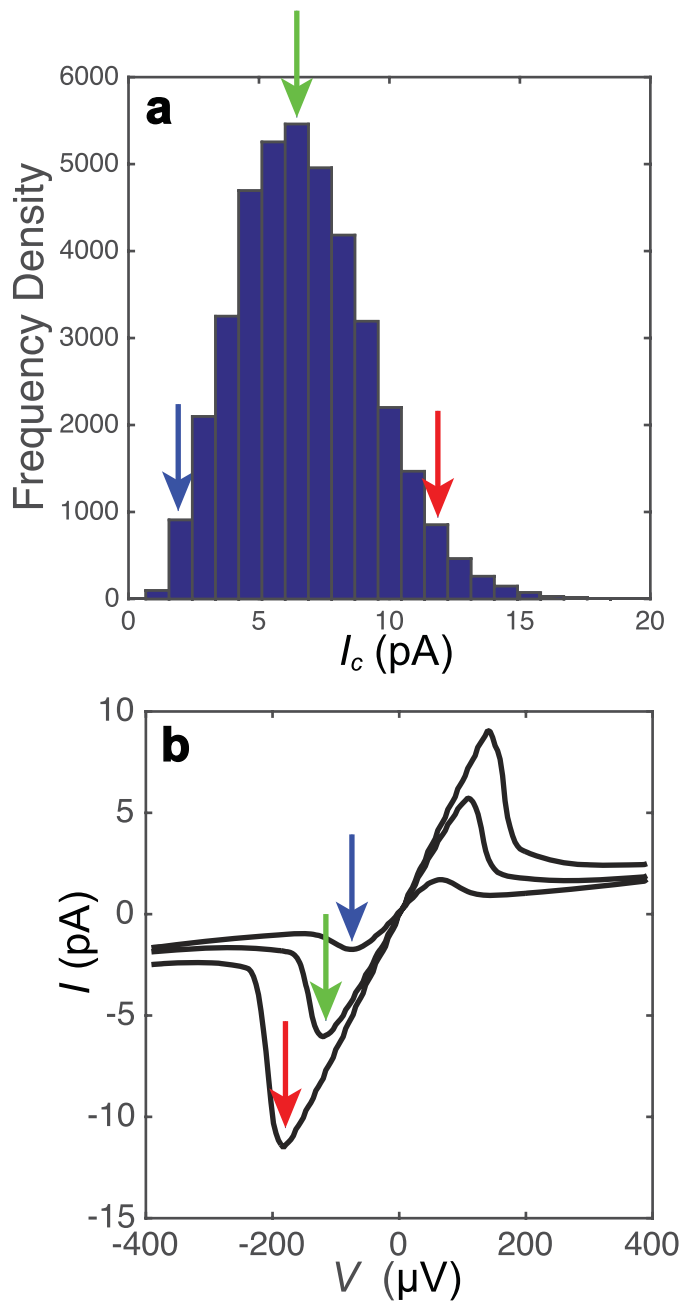


Extended Data Figure 2 | Spectroscopic/topographic data from two distinct BSCCO tips. Typical dI/dV spectrum, topography and magnitude of its Fourier transform $\tilde{T}(\mathbf{q})$ measured with two completely distinct BSCCO tips on two different BSCCO samples: tip 1 (**a**, **c** and **e**) and tip 2 (**b**, **d** and **f**).

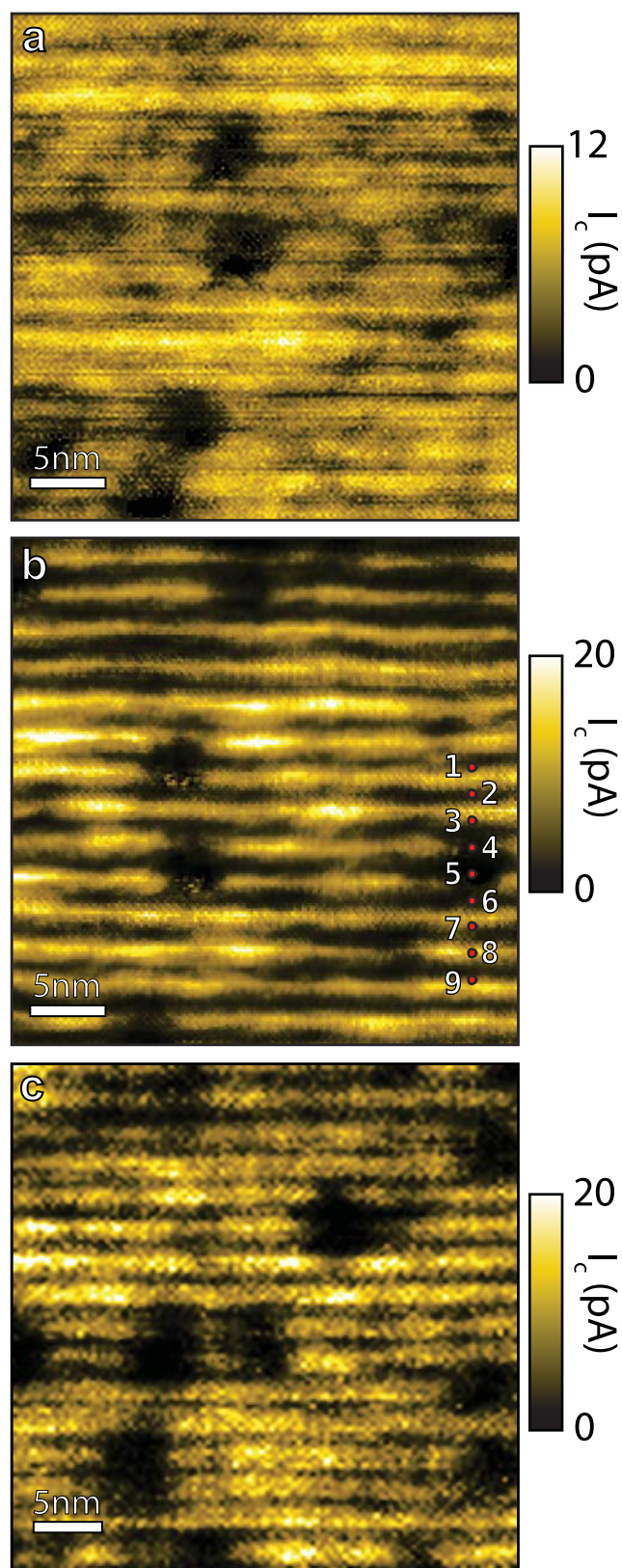


Extended Data Figure 3 | Simulation of topography with BSCCO nanoflake tip. **a**, Surface topography $T(r)$ of BSCCO sample; image obtained with a conventional metallic tungsten tip. **b**, Magnitude of Fourier transform (FT) of **a**. **c**, Model for BSCCO nanoflake tip that is aligned with the BSCCO surface in **a**. **d**, Model for BSCCO nanoflake tip that is misaligned with the BSCCO surface in **a**. **e**, Convolution of BSCCO surface image in **a** and aligned BSCCO nanoflake tip in **c**.

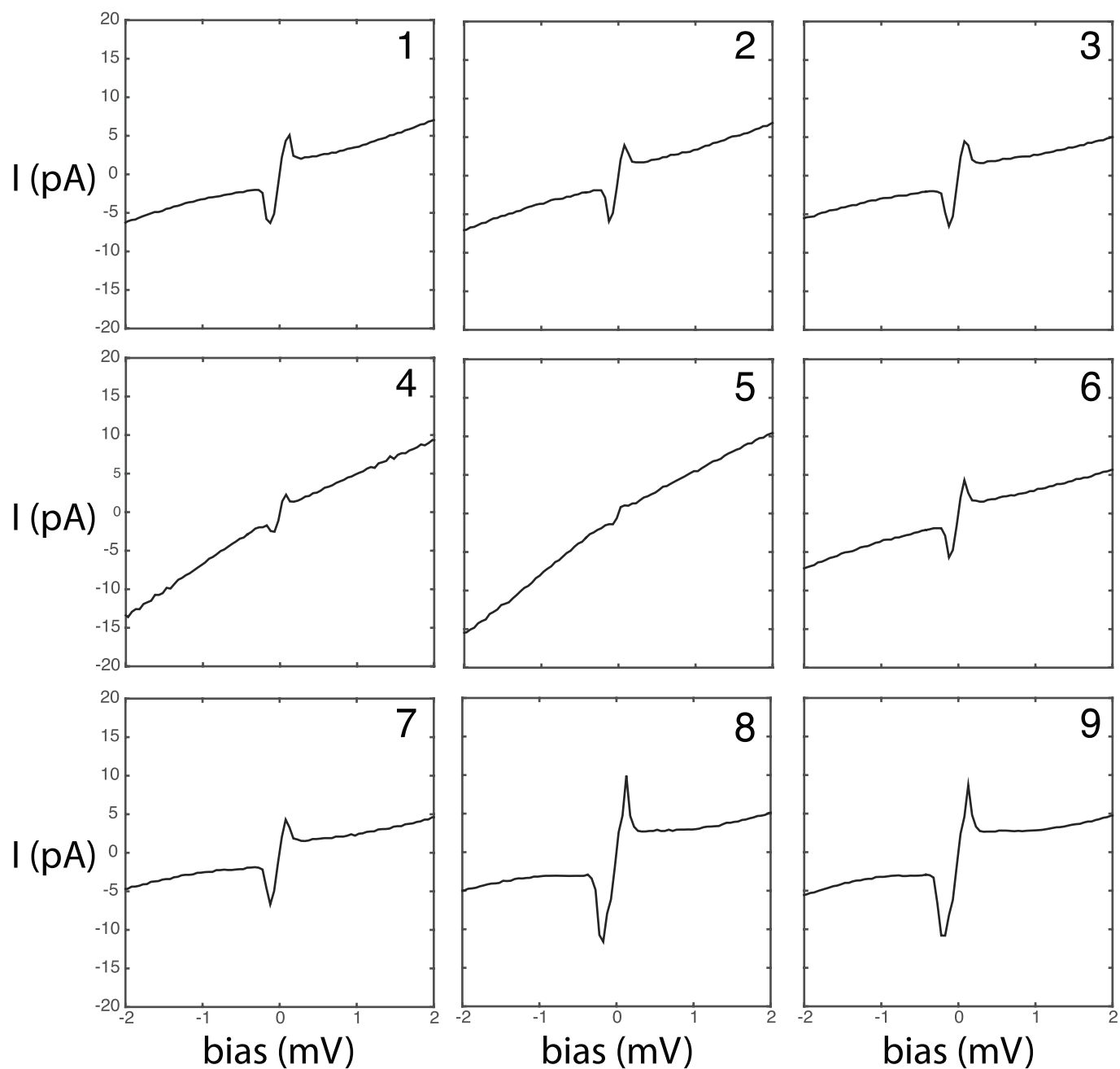
Inset shows resultant Fourier Transform with Bragg peaks in the corners and supermodulation peaks. **f**, Convolution of BSCCO surface image in **a** and misaligned BSCCO nanoflake tip in **d**. Inset shows resulting Fourier transform with many additional broad peaks caused by moiré pattern effects. Comparison to **e** demonstrates that the tip used in the studies reported here was well aligned with the sample.



Extended Data Figure 4 | Spatial variation of HTS tip-sample Josephson junction $I(V)$. **a**, A histogram of all I_c values measured at different locations in the field of view shown in Fig. 3b. The characteristic I_c values associated with the three spectra in **b** are indicated by the coloured arrows. **b**, Three $I(V)$ formed by averaging the constituent $I(V)$ from the bins indicated by the coloured arrows in **a**. They demonstrate the variation in the measured $I(V)$ characteristic that occurs at different locations in the field of view shown in Fig. 3b. The three I_c values are indicated by coloured arrows.

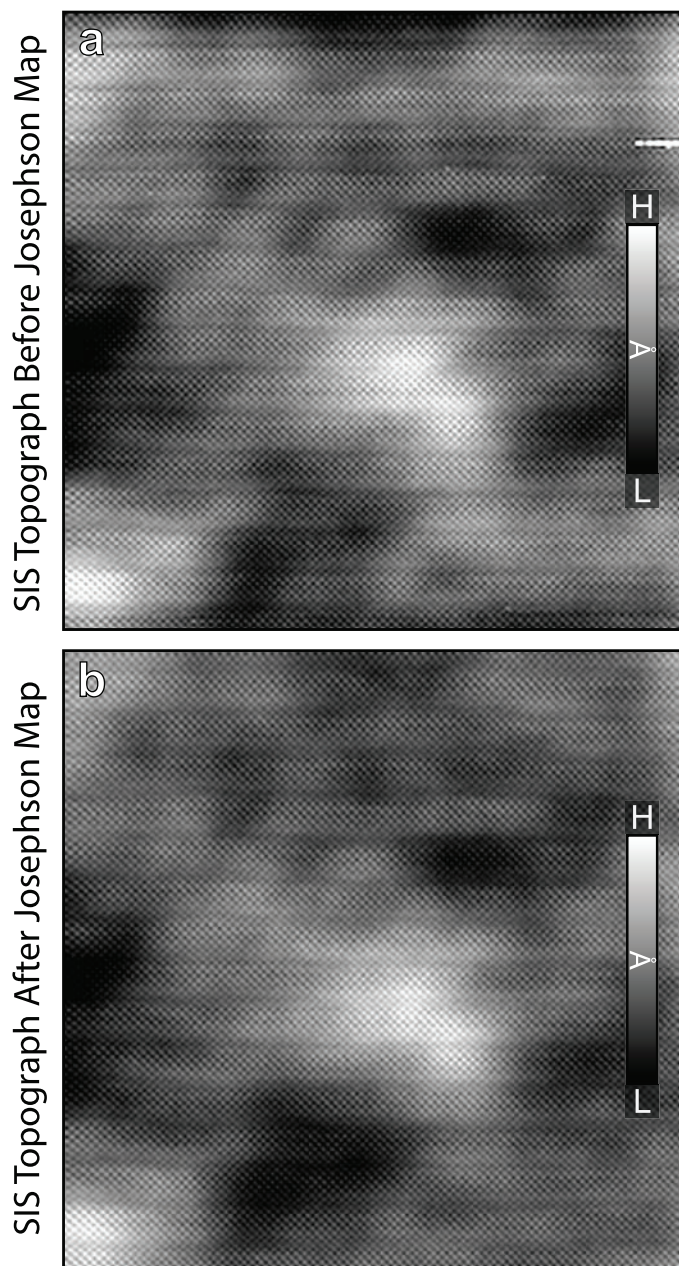


Extended Data Figure 5 | Repeatable utility of BSCCO nanoflake tips for $I_c(r)$ mapping. Three $I_c(r)$ images **a**, **b** and **c** measured with the same BSCCO nanoflake tip, at different times (separated by many days) and using different Josephson junction parameters, but closely related in fields of view. Repeatability and fidelity of our $I_c(r)$ imaging by SJTM is evident.

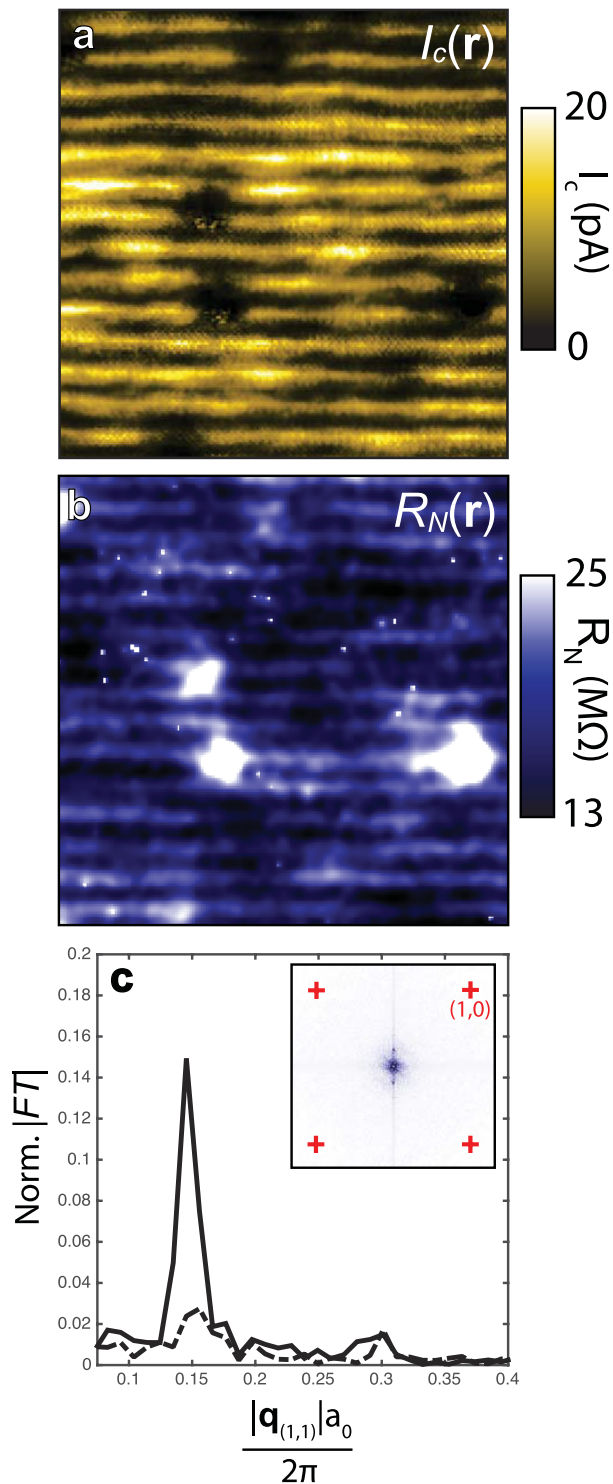


Extended Data Figure 6 | Sequence of measured $I_c(r)$ along line in Extended Data Fig. 5b. Nine individual $I(V)$ spectra measured at the locations indicated by numbers 1–9 in Extended Data Fig. 5b. In each case the transition from the SIS resistive branch to the phase diffusion

Josephson tunnelling branch is evident. Moreover, as the sequence passes through the site of a Zn atom, the value of I_c diminishes by $\sim 95\%$ from its maximum, as expected from muon spin rotation experiments. Bias, V_B .

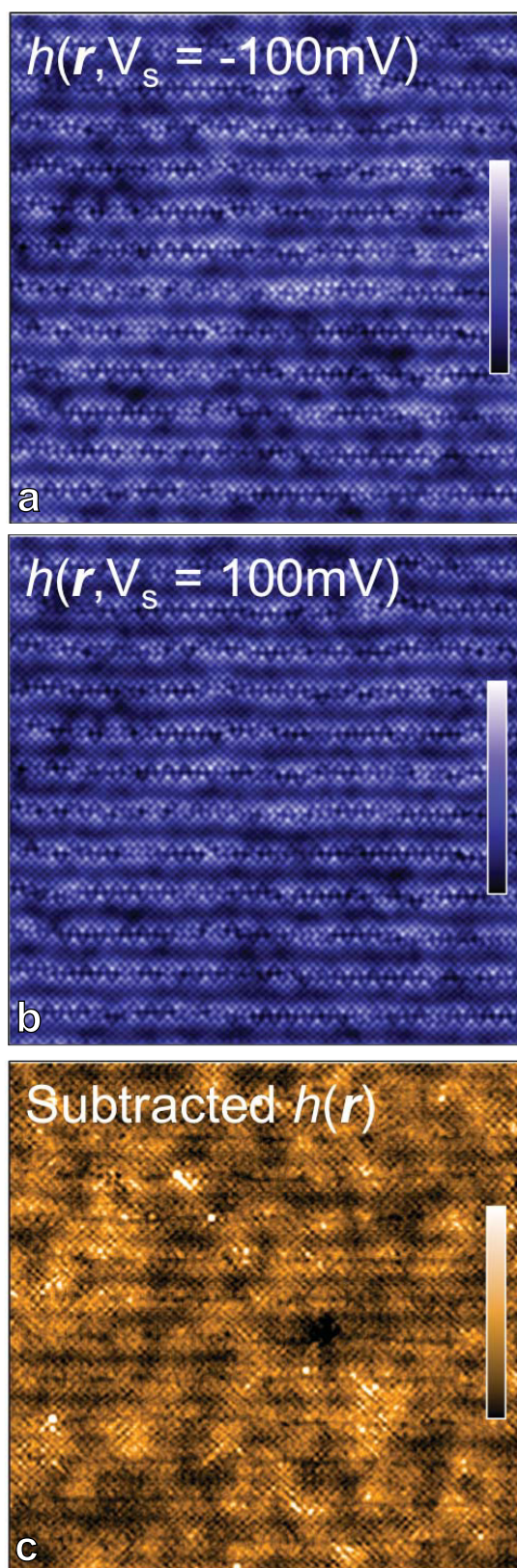


Extended Data Figure 7 | Before and after topographic images bracketing $I_c(r)$ map. **a**, Topograph taken with BSCCO tip before typical $I_c(r)$ SJTM map. **b**, Topograph taken with same BSCCO tip after the same $I_c(r)$ SJTM map as **a**. Comparison of **a** and **b** shows that tip and surface are very well preserved in our SJTM protocol.

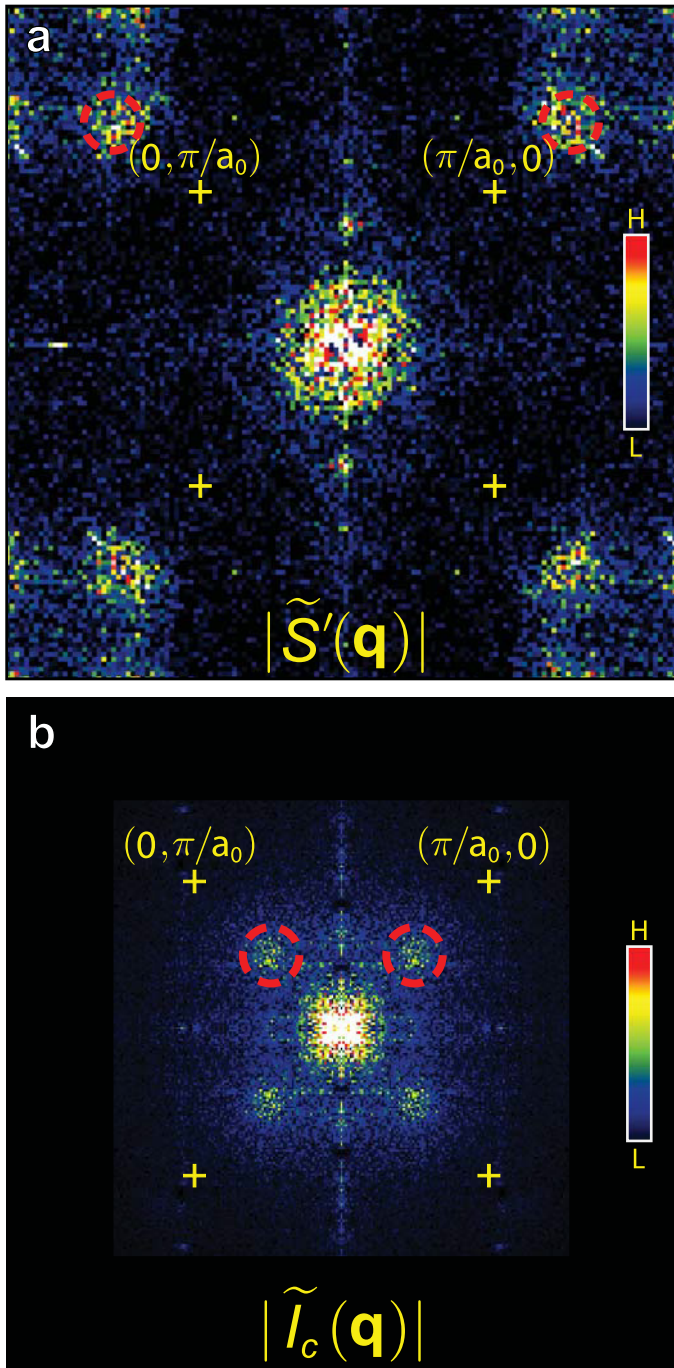


Extended Data Figure 8 | Comparison of modulations in $I_c(\mathbf{r})$ and $R_N(\mathbf{r})$.

a, A typical measured $I_c(\mathbf{r})$ image of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+x}$ with the crystal supermodulation effect retained and apparent as strong spatial modulations in I_c along the vertical axis. **b**, Measured $R_N(\mathbf{r})$ image simultaneous with **a**, with the crystal supermodulation effect retained. The spatial modulations in $R_N(\mathbf{r})$ along the vertical axis are greatly diminished in relative amplitude compared to $I_c(\mathbf{r})$ modulations in **a**. **c**, Inset shows $|\tilde{R}_N(\mathbf{q})|$, the magnitude of the Fourier transform of $R_N(\mathbf{r})$ from **b**. Plotting the simultaneously measured Fourier amplitudes of $\tilde{I}_c(\mathbf{q})$ and $\tilde{R}_N(\mathbf{q})$ along the $(1, 1)$ direction passing through the wavevector of the supermodulation \mathbf{Q}_{SM} shows that modulations in $\tilde{R}_N(\mathbf{q})$ are negligible. Therefore the predominant effect in the $I_c(\mathbf{r})R_N(\mathbf{r})$ studied through this work is caused by the $I_c(\mathbf{r})$ variations, coming from the superfluid density variations $\rho_s(\mathbf{r}) \propto I_J^2(\mathbf{r}) \propto I_c(\mathbf{r})$ of the condensate in the sample.



Extended Data Figure 9 | d -symmetry density wave $\tilde{D}(q)$ from topography. **a**, High-resolution topographic image of typical BiO surface at the same hole-density as the $I_c(r)$ studies, measured at $V = 100 \text{ meV}$. **b**, High-resolution topographic image of identical (registered for every atom within about 10 picometres) BiO surface as measured at $V = -100 \text{ meV}$. **c**, Difference between **a** and **b**; a CDW exhibits its signature logarithmically in such an image and therefore it can be used to detect the d -symmetry form factor density wave, as in Fig. 4d of the main text.



Extended Data Figure 10 | Absence CDW setup effect in $I_c(\mathbf{r})$.

a, $\tilde{T}(\mathbf{q})$, the Fourier transform magnitude of the sublattice-resolved image $O_x(\mathbf{r}) + O_y(\mathbf{r})$ derived from $\delta T(\mathbf{r})$, the difference between the two unprocessed topographic images $T(\mathbf{r}, \pm 100 \text{ meV})$ in Extended Data Fig. 9. We see directly that the actual modulations in topography due to the density of states modulations from the CDW occur at wavevectors $(1 \pm 0.22, 0)2\pi/a_0$ and $(0, 1 \pm 0.22)2\pi/a_0$ (dashed circles), as has been reported extensively in the past. These only become detectable at the actual CDW wavevector $\mathbf{Q}_C = (0.22, 0)2\pi/a_0$ and $(0, 0.22)2\pi/a_0$ when one uses a measure of d -symmetry form factor modulations: $\tilde{D}(\mathbf{q}) = \tilde{O}_x(\mathbf{q}) - \tilde{O}_y(\mathbf{q})$, as shown in Fig. 4d. Because the physically real modulations in topography and conductance imaging therefore occur at $\mathbf{Q} = (0.78, 0)2\pi/a_0$ and $(0, 0.78)2\pi/a_0$ (dashed circles), it is impossible for them to produce, through a 'setup effect', spurious $I_c(\mathbf{r})$ modulations at the PDW wavevector $\mathbf{Q}_P \approx (0.25, 0)2\pi/a_0$ and $(0, 0.25)2\pi/a_0$, as indicated by dashed circles in **b**.

b, The measured q -space structure $\tilde{l}_c(\mathbf{q})$ (which samples all sublattices in the conventional form $\tilde{O}_x(\mathbf{q}) + \tilde{O}_y(\mathbf{q}) + \tilde{C}u(\mathbf{q})$). The PDW maxima occur at $\mathbf{Q}_P \approx (0.25, 0)2\pi/a_0$ and $(0, 0.25)2\pi/a_0$.

A pressure-amplifying framework material with negative gas adsorption transitions

Simon Krause^{1*}, Volodymyr Bon^{1*}, Irena Senkowska¹, Ulrich Stoeck^{1†}, Dirk Wallacher², Daniel M. Többsen³, Stefan Zander³, Renjith S. Pillai⁴, Guillaume Maurin⁴, François-Xavier Coudert⁵ & Stefan Kaskel¹

Adsorption-based phenomena are important in gas separations^{1,2}, such as the treatment of greenhouse-gas³ and toxic-gas⁴ pollutants, and in water-adsorption-based heat pumps⁵ for solar cooling systems. The ability to tune the pore size, shape and functionality of crystalline porous coordination polymers—or metal–organic frameworks (MOFs)—has made them attractive materials for such adsorption-based applications^{3,6–8}. The flexibility and guest-molecule-dependent response^{9,10} of MOFs give rise to unexpected and often desirable adsorption phenomena^{11–14}. Common to all isothermal gas adsorption phenomena, however, is increased gas uptake with increased pressure. Here we report adsorption transitions in the isotherms of a MOF (DUT-49) that exhibits a negative gas adsorption; that is, spontaneous desorption of gas (methane and *n*-butane) occurs during pressure increase in a defined temperature and pressure range. A combination of *in situ* powder X-ray diffraction, gas adsorption experiments and simulations shows that this adsorption behaviour is controlled by a sudden hysteretic structural deformation and pore contraction of the MOF, which releases guest molecules. These findings may enable technologies using frameworks capable of negative gas adsorption for pressure amplification in micro- and macroscopic system engineering. Negative gas adsorption extends the series of counterintuitive phenomena such as negative thermal expansion^{15,16} and negative refractive indices¹⁷ and may be interpreted as an adsorptive analogue of force-amplifying negative compressibility transitions proposed for metamaterials¹⁸.

In MOFs, guest-molecule-assisted phenomena, such as ligand flip¹⁹, gating²⁰, and breathing^{21,22} of the framework lead to adsorption isotherms with an S-shape, wide hysteresis and multiple steps²³, which give rise to previously unattainable potential applications¹⁰. As a general rule of thermodynamics, at constant temperature the absolute amount of substance in the adsorbed phase (n_{ads}) always increases with increasing pressure of the adsorptive (p), resulting in a positive slope (see equation (1)) in a single-component adsorption isotherm (Fig. 1a)²⁴:

$$(dn_{\text{ads}}/dp) \geq 0 \quad (1)$$

However, here we describe a material with a negative gas adsorption (NGA) step, causing an unexpected pressure amplification triggered by a sudden structural transition of a nanoporous framework. We show that this phenomenon is observed upon adsorption of various gas molecules at characteristic temperature and pressure in DUT-49 (Dresden University of Technology No. 49), a framework with very high porosity. Initially, DUT-49 was discovered as a high-performance methane (CH₄) storage material for natural gas vehicles²⁵. Its extremely high excess methane storage capacity of 308 g kg^{−1} at 110 bar and 298 K motivated us to elucidate the adsorption mechanism at lower temperatures.

However, a methane adsorption isotherm recorded at 111 K revealed a peculiar drop that was originally regarded as an instrumental failure (Fig. 1b). Repeated isotherm measurements at 111 K gave evidence of a perfectly reproducible negative adsorption step around 10 kPa, strongly deviating from the previously reported adsorption profile at 298 K (up to 110 bar). With a drop of $\Delta n_{\text{NGA}} = -8.62 \text{ mmol g}^{-1}$ the NGA step alone surpasses the adsorption capacity of commercially available nanoparticles and even of some zeolites²⁶. Thus in volumetric adsorption experiments it causes a massive response and pressure amplification of $\Delta p_{\text{NGA}} = 2.27 \text{ kPa}$ in the measuring cell (Fig. 1f), in contrast to what happens in any rigid microporous solids (Fig. 1e). This corresponds to a specific Δp_{NGA} for methane adsorption at 111 K of up to 493 kPa g^{−1} (Supplementary Table 1).

Additional adsorption experiments using *n*-butane at 298 K lead to a distinct NGA step at 30 kPa (Fig. 1c) with a kinetic profile (Fig. 1g) similar to that of methane (111 K) while isotherms of N₂ at 77 K (Supplementary Fig. 5) show a typical 'type I' slope, illustrating the subtle role of the gaseous guest molecule (N₂ in contrast to CH₄ and C₄H₁₀) and its specific interaction stimulating NGA in DUT-49. Gravimetric adsorption experiments with *n*-butane at 298 K in parallel with dynamic scanning calorimetry support the results obtained from volumetric adsorption experiments and also reveal a complex combination of exothermic and endothermic processes during NGA (Fig. 1d, Supplementary Fig. 24). In particular, exposing samples of DUT-49 to *n*-butane permits us to investigate NGA at room temperature below 100 kPa, conditions that are easily attained in a standard laboratory environment. Visual monitoring of the sample during the adsorption experiment reveals macroscopic movement of the particle ensemble caused by the massive gas release during NGA, followed by a pronounced shrinkage, by about 20%, of the packed bed volume (Fig. 1h, Supplementary Video 1). From this macroscopic observation we suspected NGA to be related to massive structural transformations in the crystalline solid. Real-time parallelized adsorption/diffraction/EXAFS (extended X-ray absorption fine structure) experiments carried out at the BESSY II light source at the Helmholtz-Zentrum Berlin using an improved setup²⁷ allowed us to elucidate the structural transformations *in situ*.

Synchrotron powder X-ray diffraction (PXRD) data collected during methane adsorption at 111 K indicate two distinct structural transitions of the open-pore structure DUT-49_{op} at 10 kPa and 45–65 kPa correlating with the steps in the isotherm (Fig. 2a and b). DUT-49_{op} is an assembly of cuboctahedral metal–organic polyhedra (MOPs) formed by Cu₂ paddle-wheel units and 9*H*-carbazole-3,6-dicarboxylates connected by biphenylene units resulting in face-centred cubic analogue arrangement of MOPs (a framework with *fcu* topology)²⁵. Wide open voids (10 Å, 17 Å and 24 Å) originate from an ordered arrangement of MOPs, an expanded analogue of a cubic close packing of atoms

¹Department of Inorganic Chemistry, Technische Universität Dresden, Bergstrasse 66, 01062 Dresden, Germany. ²Department Sample Environments, Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany. ³Department Structure and Dynamics of Energy Materials, Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin, Germany. ⁴Institut Charles Gerhardt Montpellier UMR 5253 CNRS UM ENSCM, Université Montpellier, Place E. Bataillon, 34095 Montpellier cedex 05, France. ⁵Chimie ParisTech, PSL Research University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France. [†]Present address: Institute for Complex Materials, Leibniz Institute for Solid State and Materials Research (IFW) Dresden, Helmholtzstraße 20, D-01069 Dresden, Germany.

*These authors contributed equally to this work.

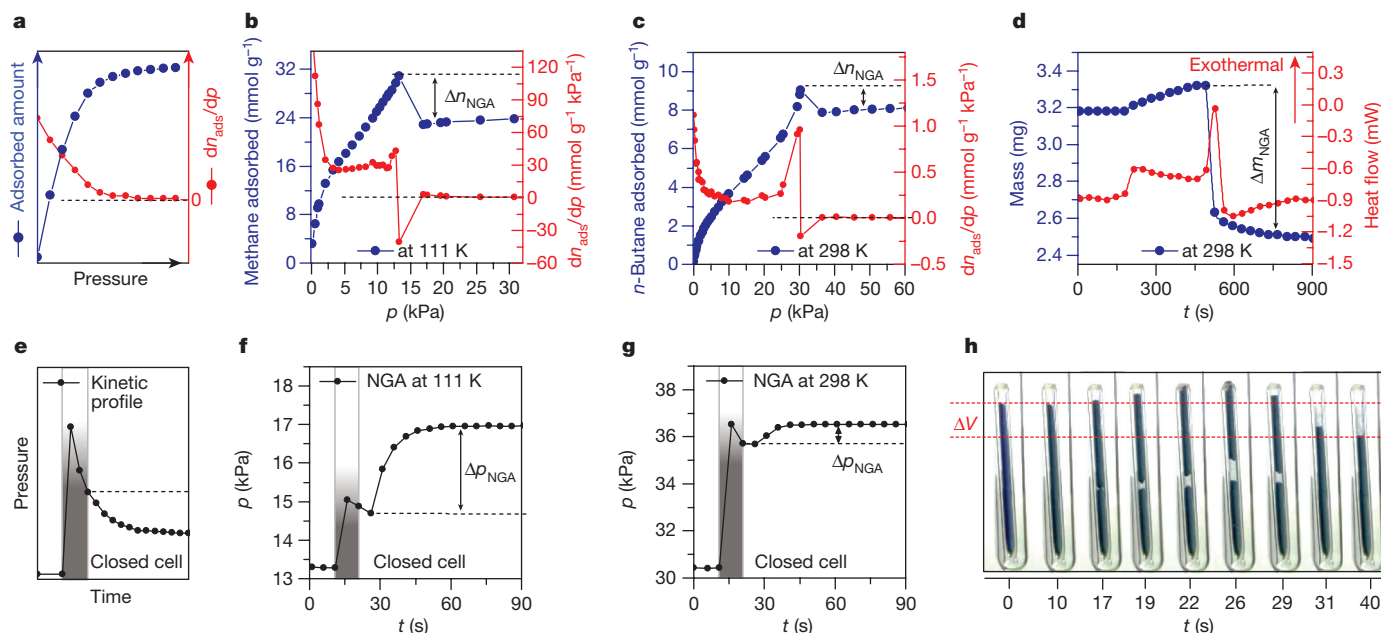


Figure 1 | Gas adsorption isotherms, pressure and kinetic profiles. **a**, Typical profile of type I adsorption isotherms for a non-breathing microporous solid and corresponding dn_{ads}/dp (ref. 24). **b**, **c**, Adsorption isotherms (blue) and corresponding derivatives (red) of DUT-49 for methane at 111 K (**b**) and *n*-butane at 298 K (**c**). Δn_{NGA} represents the NGA step in the isotherm. **d**, Gravimetric adsorption in parallel with dynamic scanning calorimetry, Δm_{NGA} represents the mass of the gas

released during NGA. **e–g**, Corresponding kinetic pressure profiles for the typical adsorption behaviour of a non-breathing microporous solid (**e**), and for DUT-49 observed during the NGA step for methane at 111 K (**f**) and *n*-butane at 298 K (**g**); grey areas indicate the gas dosing interval. **h**, Snapshots of Supplementary Video 1 recorded during the *n*-butane adsorption steps at 298 K (**c**, **g**) (the capillary diameter is 1 mm and the sample bed is 5 cm in height at time $t = 0$ s).

but with huge tetrahedral and octahedral voids (Fig. 3a–c). During NGA the highly porous framework undergoes an enormous contraction of the unit cell volume from $V = 100,072 \text{ \AA}^3$ (calculated density $\rho_{\text{calc}} = 0.287 \text{ g cm}^{-3}$) to a highly squeezed intermediate, the contracted-pore structure DUT-49cp with a unit cell volume of only $47,282 \text{ \AA}^3$ ($\rho_{\text{calc}} = 0.661 \text{ g cm}^{-3}$) (Supplementary Fig. 34), corresponding to a unit cell change in volume of more than $50,000 \text{ \AA}^3$ (Fig. 2c). This initial transformation is reversible: as the pores become completely filled, expansion back to the original DUT-49op is observed (Supplementary Fig. 35).

From a structural point of view the contraction originates from concerted rotation of the MOPs around the threefold [111] axis leading to a severe deformation of the ligand (Fig. 3e, Supplementary Video 2). Yet the connectivity of the framework is preserved during the transformation and the porosity of the MOPs (pore diameter $10.1\text{--}9.9 \text{ \AA}$) is only slightly affected. EXAFS spectra collected *in situ* during methane adsorption at 111 K confirm the unaltered coordination environment of Cu^{2+} (Supplementary Figs 38 and 39, Supplementary Table 8). In contrast, the pore diameter of the tetrahedral voids is reduced from 16.9 \AA to 10.6 \AA and the mesoporous octahedral voids deform even more, from

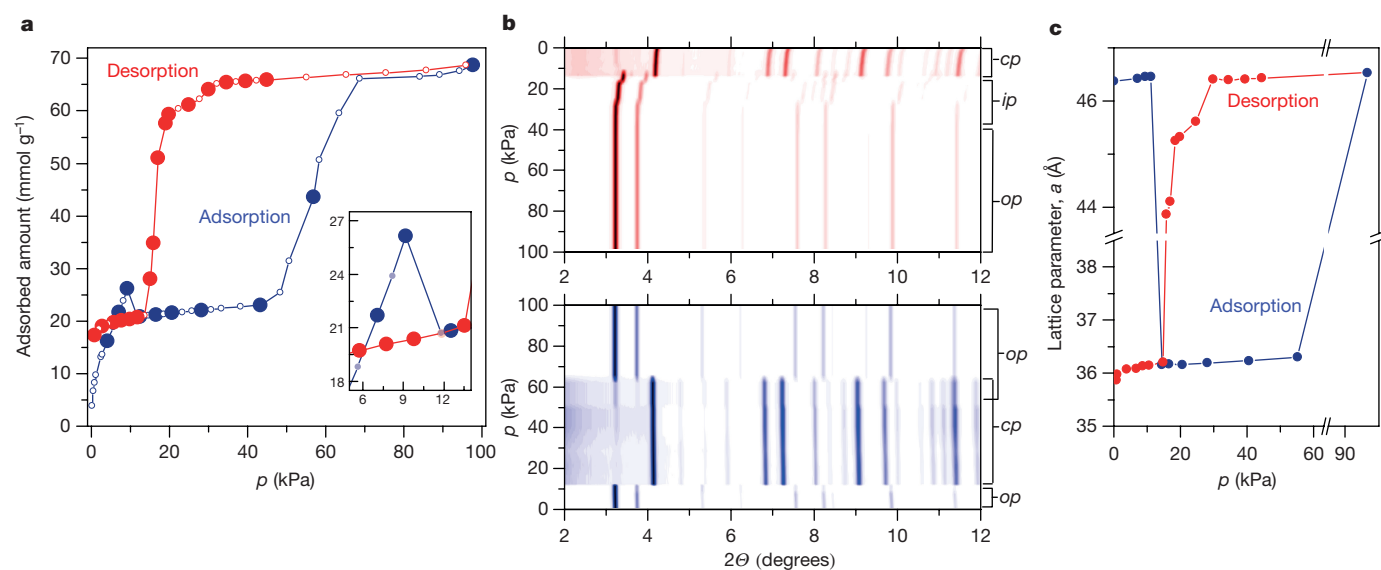


Figure 2 | In situ PXRD during methane adsorption at 111 K. **a**, Methane adsorption and desorption isotherms at 111 K collected *in situ*. The points corresponding to collected PXRD patterns are represented as solid symbols; the inset shows an enlarged NGA region of the isotherm. **b**, Contour plot of the PXRD pattern evolution ($\lambda = 1.5406 \text{ \AA}$) during

adsorption (blue) and desorption (red). *op* is the open-pore phase (cubic, space group $F23$, lattice parameter $a = 46.427(4) \text{ \AA}$), *cp* is the contracted-pore phase (cubic, $Pa\bar{3}$, $a = 36.1603(2) \text{ \AA}$), and *ip* is the intermediate pore phase (cubic, $Pa\bar{3}$, $a = 45.542(2) \text{ \AA}$). **c**, Lattice parameter evolution during adsorption (blue) and desorption (red) of methane at 111 K.

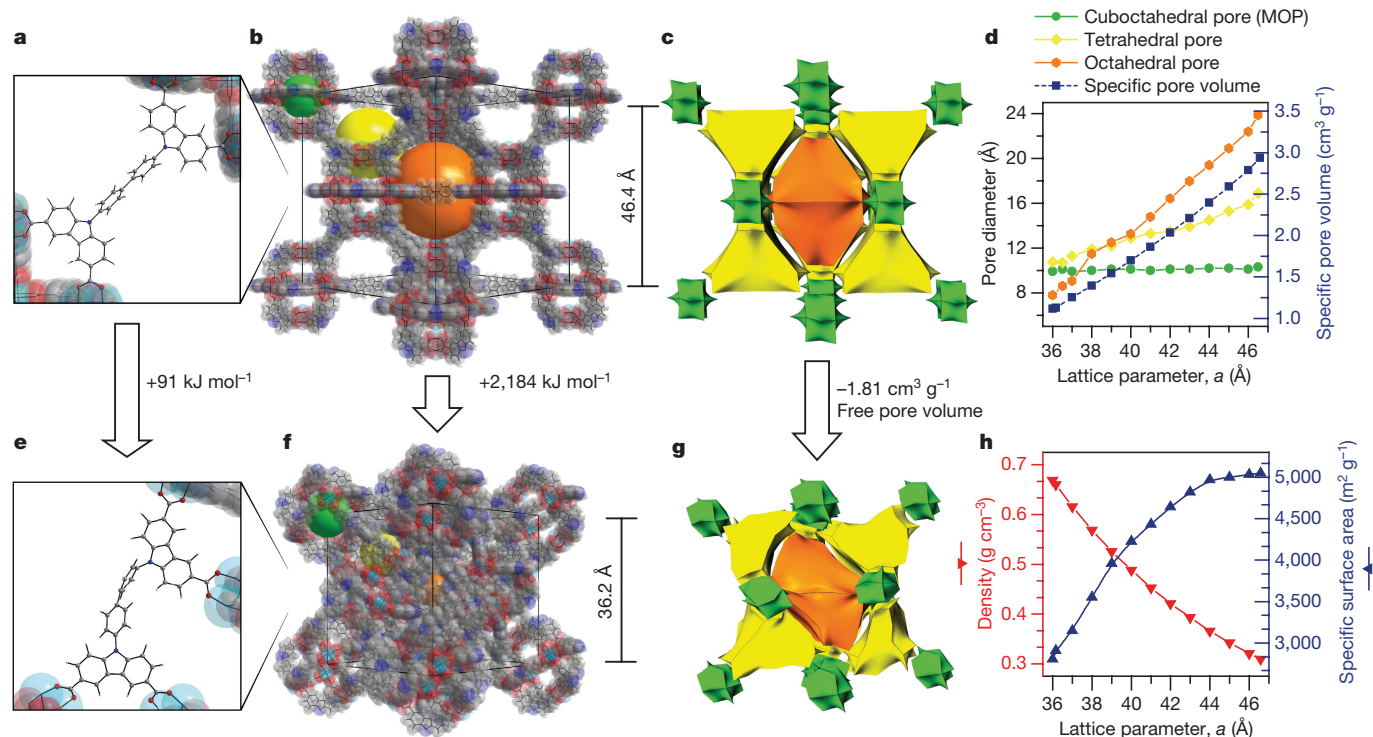


Figure 3 | Structure of DUT-49op and DUT-49cp. a, e, Conformation of 9,9'-(1,1'-biphenyl)-4,4'-diylbis(9H-carbazole-3,6-dicarboxylate) in DUT-49op (a) and DUT-49cp (e). b, f, View of the crystal structure along [110] for DUT-49op (simulated accessible surface area 5,049 m² g⁻¹, free pore volume 2.94 cm³ g⁻¹) (b) and DUT-49cp (simulated accessible surface area 2,908 m² g⁻¹, free pore volume 1.13 cm³ g⁻¹) (f). Atoms are coloured as follows: C (dark grey), N (blue), O (red), Cu (cyan) and H (light grey). Guest molecules are omitted for clarity. c, g, Pore representation of

DUT-49op (c) and DUT-49cp (g) along the [110] axis as natural tilings of the **nbo** topology net (the copper paddle wheels forming the (4,4)-connected network structure of NbO). Pores are coloured as follows: cuboctahedral MOP cavity (green), tetrahedral cavity (yellow) and octahedral cavity (orange). d, Evolution of pore sizes and free pore volume during contraction from DUT-49op to DUT-49cp. h, Evolution of density and accessible geometric surface area. Details of the simulation of intermediate structures, free pore volume, accessible surface area and pore sizes are given in Supplementary Fig. 40.

23.9 Å to 7.8 Å, resulting in an enormous overall pore volume reduction of 61% (see also visualization in Supplementary Video 2).

During desorption of methane at 111 K, two gradual phase transitions from DUT-49op to the intermediate-pore structure DUT-49ip ($V = 94,457 \text{ Å}^3$, $a = 45.54 \text{ Å}$, $\rho = 0.332 \text{ g cm}^{-3}$) in the pressure range 30–20 kPa and from DUT-49ip to DUT-49cp at pressures lower than 20 kPa were observed, resulting in a wide hysteresis. Remarkably, during desorption at 111 K the framework does not switch back to the open form even at very low pressures (Supplementary Fig. 6). However, transformation to DUT-49op can be achieved by heating at a constant methane pressure of 100 kPa from 111 K to 298 K followed by evacuation (Supplementary Figs 7 and 9). *In situ* PXRD/adsorption experiments of *n*-butane at 298 K show DUT-49op to DUT-49cp transitions similar to those observed for methane adsorption at 111 K (Supplementary Fig. 32).

We used grand canonical Monte Carlo simulations performed for methane at 111 K in both DUT-49op and DUT-49cp to examine the microscopic adsorption mechanism. The experimental adsorption isotherm is precisely reproduced by simulating two adsorption regimes, starting with pore filling of DUT-49op up to 10 kPa, but continued adsorption in the contracted DUT-49cp (Fig. 4b, analogous observation for *n*-butane at 298 K; Supplementary Fig. 41). Methane initially occupies the smaller MOP cavities before gradually travelling over the tetrahedral and octahedral cages of DUT-49op with increasing pressure. After contraction above 11 kPa, the entire available space of the pores is occupied by the molecules (Fig. 4c). Indeed, this scenario strongly suggests that the filling of the mesopores triggers the structure contraction responsible for NGA.

Further analysis by quantum chemistry calculations using density functional theory showed that guest-molecule-free DUT-49cp is energetically substantially destabilized ($\Delta E = 91 \text{ kJ mol}^{-1}$ per linker molecule) compared to DUT-49op owing to the highly strained organic ligand

(Fig. 3a, e). Upon adsorption, this energy difference is overcome by the higher affinity of methane for the contracted-pore structure because filling the much smaller pores of DUT-49cp offers energetic benefits. Given the adsorption enthalpies of DUT-49cp and DUT-49op ($|\Delta H_{\text{ads}}| \approx 17 \text{ kJ mol}^{-1}$ and 10 kJ mol^{-1} , respectively) (Supplementary Fig. 44 and Supplementary Table 13), the difference in adsorption enthalpy in the transition region, $\Delta \Delta H_{\text{ads}} = n_{\text{ads}}(\text{cp}) \cdot |\Delta H_{\text{ads}}(\text{op}) - \Delta H_{\text{ads}}(\text{cp})| \approx 130 \text{ kJ mol}^{-1}$ (Supplementary Equation (8)), is comparable in magnitude to ΔE , which suggests that adsorption compensates for the energy penalty of the structural transition (Supplementary Table 15). Analysis of the elastic properties of DUT-49op completes this picture: although the bulk modulus of 8.7 GPa is in the usual range for MOFs²⁸, the minimal shear modulus of the structure is relatively low at 1.5 GPa, with a strong nonlinearity showing a soft deformation mode that can be triggered by adsorption.

To summarize, in NGA a highly overloaded metastable state is created during gas adsorption in a hierarchical pore structure resulting in a spontaneous gas release and structural relaxation when the energy barrier is overcome. The result is a sudden reduction of porosity through network deformation triggered by adsorption-induced stress and the subsequent release of previously adsorbed gas²⁹. Extended equilibration times of up to several hours (Supplementary Fig. 14) show that this phenomenon is not caused by inadequate equilibration of the system and indicate the presence of long-lived metastable states, possibly due to the vast extent of deformation and crossover of pore sizes (Fig. 3d), compared with the well known, purely microporous 'breathing' (that is, flexible or switchable) systems such as MIL-53²¹. This implies the internal reshuffling of molecules in DUT-49 from rigid MOP pores into contracted octahedral voids. The temperature dependence of NGA supports this thesis: the adsorption of methane at 91 K, 111 K, 101 K and 121 K (Fig. 4a, d) and of *n*-butane at 273–308 K (Supplementary Fig. 10) even shows a slight increase in $|\Delta n_{\text{NGA}}|$ with

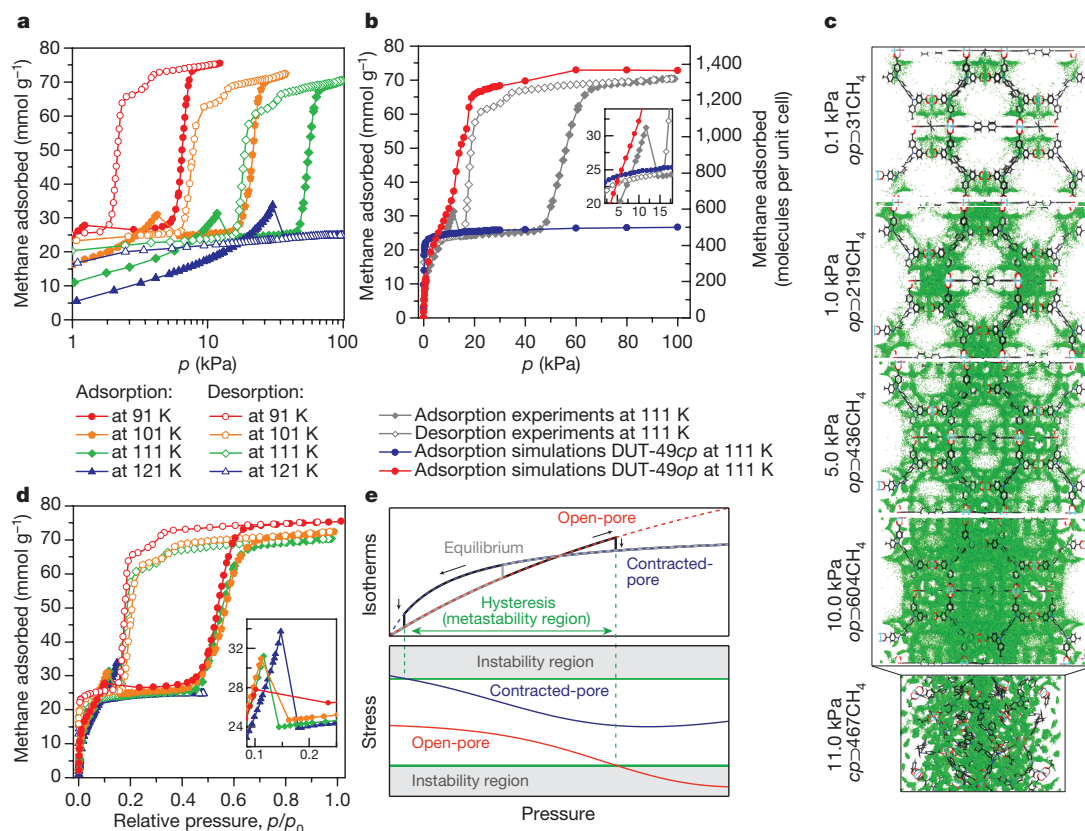


Figure 4 | Temperature-dependent methane adsorption and simulations. **a, d**, Methane adsorption isotherms at different temperatures. The inset in **d** shows an enlarged NGA region of the isotherm. **b**, Experimental (grey) and simulated methane sorption isotherm at 111 K based on the crystal structures of DUT-49op (red) and DUT-49cp (blue); the inset shows an enlarged NGA region of the isotherm. **c**, Maps of the occupied positions of methane (green) in 1,000 equilibrated

frames for a given pressure for DUT-49op and at the transition pressure (11 kPa) of DUT-49cp (bottom frame). Atoms are coloured as in Fig. 3. **e**, Schematic representation of stress-based model for adsorption-induced structural transformation²⁹: equilibrium transition (grey line), metastable transition (black arrows), open-pore phase (red line), contracted-pore phase (blue line); arrows indicate the transition observed for DUT-49.

temperature up to 9.87 mmol of methane per gram of MOF at 121 K. However, the pore contraction mechanism is maintained, as reflected by the almost identical relative pressure ($p/p_0 = 0.1$ – 0.15) (where p_0 is the saturation pressure of the bulk fluid) at which NGA is observed (Supplementary Figs 27 and 30). This suggests that filling of the mesopores induces the contraction (Fig. 4a, c).

To our knowledge, the phenomenon of NGA in MOFs has not been reported before. The rare metastable cation complexes in zeolite materials could be considered to have analogous properties³⁰. Besides the fundamental interest of this phenomenon, there are many possible technological applications. These porous solids could enable the design of threshold-sensitive micropneumatic devices, stimuli-responsive self-propelling systems or mechanical actuators responding selectively to changes in the environment, and capable of transforming a large amount of latent strain into pressure.

Received 3 October 2015; accepted 16 February 2016.

Published online 6 April 2016.

- de Vos, R. M. & Verweij, H. High-selectivity, high-flux silica membranes for gas separation. *Science* **279**, 1710–1711 (1998).
- Nenoff, T. M. Hydrogen purification: MOF membranes put to the test. *Nature Chem.* **7**, 377–378 (2015).
- McDonald, T. M. *et al.* Cooperative insertion of CO₂ in diamine-appended metal-organic frameworks. *Nature* **519**, 303–308 (2015).
- Barea, E., Montoro, C. & Navarro, J. A. R. Toxic gas removal—metal-organic frameworks for the capture and degradation of toxic gases and vapours. *Chem. Soc. Rev.* **43**, 5419–5430 (2014).
- Furukawa, H. *et al.* Water adsorption in porous metal-organic frameworks and related materials. *J. Am. Chem. Soc.* **136**, 4369–4381 (2014).
- Kitagawa, S., Kitaura, R. & Noro, S.-i. Functional porous coordination polymers. *Angew. Chem. Int. Ed.* **43**, 2334–2375 (2004).

- Furukawa, H., Cordova, K. E., O’Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal-organic frameworks. *Science* **341**, <http://dx.doi.org/10.1126/science.1230444> (2013).
- Nugent, P. *et al.* Porous materials with optimal adsorption thermodynamics and kinetics for CO₂ separation. *Nature* **495**, 80–84 (2013).
- Horike, S., Shimomura, S. & Kitagawa, S. Soft porous crystals. *Nature Chem.* **1**, 695–704 (2009).
- Schneemann, A. *et al.* Flexible metal-organic frameworks. *Chem. Soc. Rev.* **43**, 6062–6096 (2014).
- Matsuda, R. Materials chemistry: selectivity from flexibility. *Nature* **509**, 434–435 (2014).
- Sato, H. *et al.* Self-accelerating CO sorption in a soft nanoporous crystal. *Science* **343**, 167–170 (2014).
- Takahashi, Y. *et al.* Molecular decoding using luminescence from an entangled porous framework. *Nature Commun.* **2**, 168 (2011).
- Yang, S. *et al.* A partially interpenetrated metal-organic framework for selective hysteretic sorption of carbon dioxide. *Nature Mater.* **11**, 710–716 (2012).
- Martinek, C. & Hummel, F. A. Linear thermal expansion of three tungstates. *J. Am. Ceram. Soc.* **51**, 227–228 (1968).
- Takenaka, K. Negative thermal expansion materials: technological key for control of thermal expansion. *Sci. Technol. Adv. Mater.* **13**, 013001 (2012).
- Shelby, R. A., Smith, D. R. & Schultz, S. Experimental verification of a negative index of refraction. *Science* **292**, 77–79 (2001).
- Nicolaou, Z. G. & Motter, A. E. Mechanical metamaterials with negative compressibility transitions. *Nature Mater.* **11**, 608–613 (2012).
- Fairen-Jimenez, D. *et al.* Opening the gate: framework flexibility in ZIF-8 explored by experiments and simulations. *J. Am. Chem. Soc.* **133**, 8900–8902 (2011).
- Kondo, A. *et al.* Novel expansion/shrinkage modulation of 2D layered MOF triggered by clathrate formation with CO₂ molecules. *Nano Lett.* **6**, 2581–2584 (2006).
- Serre, C. *et al.* An explanation for the very large breathing effect of a metal-organic framework during CO₂ adsorption. *Adv. Mater.* **19**, 2246–2251 (2007).
- Férey, G. & Serre, C. Large breathing effects in three-dimensional porous hybrid matter: facts, analyses, rules and consequences. *Chem. Soc. Rev.* **38**, 1380–1399 (2009).

23. Salles, F. *et al.* Multistep N₂ breathing in the metal–organic framework Co(1,4-benzenedipyrazolate). *J. Am. Chem. Soc.* **132**, 13782–13788 (2010).
24. Thommes, M. *et al.* Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report). *Pure Appl. Chem.* **87**, 1051–1069 (2015).
25. Stoeck, U., Krause, S., Bon, V., Senkova, I. & Kaskel, S. A highly porous metal–organic framework, constructed from a cuboctahedral super-molecular building block, with exceptionally high methane uptake. *Chem. Commun.* **48**, 10841–10843 (2012).
26. Traa, Y. & Weitkamp, J. in *Handbook of Porous Solids* (eds, Schüth, F., Sing, K. S. W. & Weitkamp, J.) 1015–1057 (Wiley, 2008).
27. Bon, V. *et al.* In situ monitoring of structural changes during the adsorption on flexible porous coordination polymers by X-ray powder diffraction: instrumentation and experimental results. *Micropor. Mesopor. Mater.* **188**, 190–195 (2014).
28. Tan, J. C. & Cheetham, A. K. Mechanical properties of hybrid inorganic–organic framework materials: establishing fundamental structure–property relationships. *Chem. Soc. Rev.* **40**, 1059–1080 (2011).
29. Neimark, A. V., Coudert, F.-X., Boutin, A. & Fuchs, A. H. Stress-based model for the breathing of metal–organic frameworks. *J. Phys. Chem. Lett.* **1**, 445–449 (2010).
30. Riekert, L. Instability of sorption complexes in synthetic faujasites. *J. Phys. Chem.* **73**, 4384–4386 (1969).

Supplementary Information is available in the online version of the paper.

Acknowledgements V.B. thanks the German Federal Ministry for education and research (project BMBF number 05K130D3). S. Krause, V.B., I.S. and S. Kaskel thank the Helmholtz-Zentrum Berlin for financial support and allocation of

synchrotron radiation beam time at the KMC-2 beamline. F.-X.C. thanks GENCI (grant number x2015087069) for access to High-Performance Computing resources. G.M. thanks Institut Universitaire de France for its support. We thank U. Koch for scanning electron microscope images, as well as L. Sarkisov and A. Fuchs for discussions.

Author Contributions S. Krause and U.S. synthesized and activated the material investigated; S. Krause, V.B., I.S. and S. Kaskel conducted and interpreted volumetric and gravimetric sorption experiments. S. Krause, V.B., D.W. and D.M.T. conducted the *in situ* PXRD investigations. S. Krause, V.B., D.W. and S.Z. conducted the *in situ* EXAFS investigations. V.B. solved the crystal structure and performed structure modelling. S. Krause performed filming and animation of the structural transitions. F.-X.C. performed quantum mechanical calculations. G.M. and R.S.P. performed the grand canonical Monte Carlo simulations of adsorption. S. Krause, V.B., I.S., F.-X.C., G.M. and S. Kaskel wrote the manuscript.

Author Information The following crystal structures are available free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC-1413081, CCDC-1413083, CCDC-1413082 and CCDC-1413084, respectively: DUT-49*op* in vacuum at 111 K; DUT-49*cp*⊃432CH₄ (where ⊃ indicates the number of methane molecules per unit cell) at a relative pressure of 0.28 for methane at 111 K during adsorption; DUT-49*op*⊃1,344CH₄ at a relative pressure of 0.97 for methane at 111 K during adsorption; and DUT-49*ip*⊃1,176CH₄ at a relative pressure of 0.25 for methane at 111 K during desorption. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. Kaskel (stefan.kaskel@chemie.tu-dresden.de).

Copper-catalysed enantioselective stereodivergent synthesis of amino alcohols

Shi-Liang Shi¹, Zackary L. Wong¹ & Stephen L. Buchwald¹

The chirality, or ‘handedness’, of a biologically active molecule can alter its physiological properties. Thus it is routine procedure in the drug discovery and development process to prepare and fully characterize all possible stereoisomers of a drug candidate for biological evaluation^{1,2}. Despite many advances in asymmetric synthesis, developing general and practical strategies for obtaining all possible stereoisomers of an organic compound that has multiple contiguous stereocentres remains a challenge³. Here, we report a stereodivergent copper-based approach for the expeditious construction of amino alcohols with high levels of chemo-, regio-, diastereo- and enantioselectivity. Specifically, we synthesized these amino-alcohol products using sequential, copper-hydride-catalysed hydrosilylation and hydroamination of readily available enals and enones. This strategy provides a route to all possible stereoisomers of the amino-alcohol products, which contain up to three contiguous stereocentres. We leveraged catalyst control and stereospecificity simultaneously to attain exceptional control of the product stereochemistry. Beyond the immediate utility of this protocol, our strategy could inspire the development of methods that provide complete sets of stereoisomers for other valuable synthetic targets.

Different stereoisomers of drugs can have distinct therapeutic properties or adverse effects, because of the chiral environments provided by enzymes and receptors in biological systems. The most well known example is thalidomide, the (*R*)-enantiomer of which was an effective sedative, while the (*S*)-enantiomer caused severe teratogenic side effects and resulted in birth defects during the 1950s. Stereoisomers of drugs can also have contrasting indications, as in the case of quinine and quinidine, or even opposing biological activities (Fig. 1a, b). For these reasons, regulatory agencies require the bioactivity of all stereoisomers of pharmaceutical candidates to be evaluated during the drug discovery and development process^{1,2}. Furthermore, manufacturers must develop assays by which to determine stereochemical purity to ensure drug safety. Consequently, all stereoisomers of a molecule must be prepared for use in biological testing or as standard samples. Thus, the construction of complete stereoisomeric sets represents a practically important synthetic problem, as well as a fundamentally important research topic^{1,2}.

In the past few decades, there has been tremendous progress in the field of asymmetric synthesis, providing numerous chiral biologically active compounds with high levels of selectivity³. However, although asymmetric catalysis has allowed enantiomers of a chiral molecule to be obtained with equal ease, relatively few methods can provide a unified route that leads to all possible stereoisomers of products containing multiple contiguous stereocentres. Thus, full control of absolute and relative stereochemical configuration remains an unmet synthetic challenge³. Aside from classical techniques in asymmetric catalysis—such as using additives^{4,5}, modifying catalyst structure^{6–10}, and varying substrate-protecting groups^{11,12}—multicatalytic approaches^{13–16} have been advanced to access the full complement of stereoisomers for certain classes of compounds. For example, MacMillan and colleagues^{13,14} described the novel concept of cycle-specific amino-catalysis, in which

two chiral catalysts sequentially perform an iminium/enamine catalysis cascade to functionalize enals selectively. More recently, Carreira and colleagues^{15,16} demonstrated an elegant dual catalyst system for independently controlling the two stereocentres during the α -allylation of branched aldehydes. Nonetheless, a rapid and predictable way to access complete stereoisomeric sets of products bearing multiple stereocentres (for example, three contiguous stereocentres), using readily available precursors and based on a single catalyst system, remains underdeveloped but is highly desirable.

Optically pure amino alcohols are important structural elements that are frequently found in pharmaceutical agents and biologically active natural products (Fig. 1b)¹⁷. These compounds are also building blocks for catalysts and auxiliaries in asymmetric synthesis. We speculated that our recently disclosed catalytic systems^{18–23} for enantioselective hydroamination^{24,25} based on copper hydride (CuH) intermediates²⁶

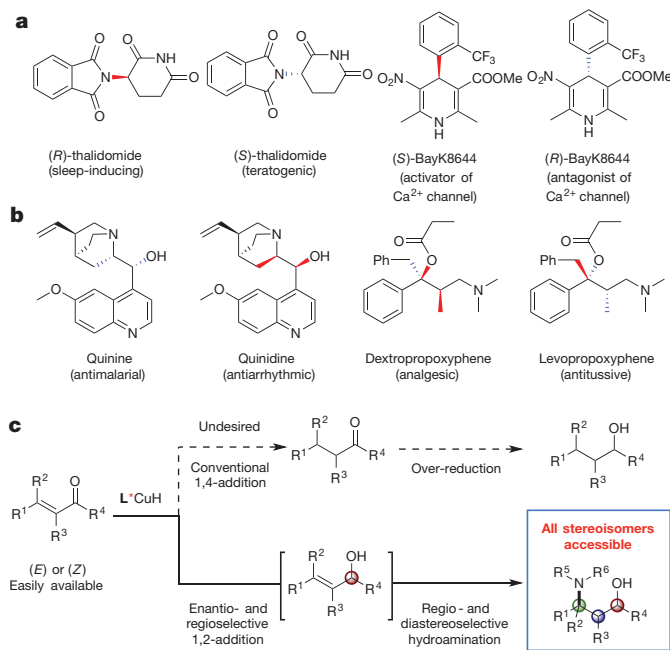


Figure 1 | Varying biological activity of some different stereoisomers, and our strategy for constructing all stereoisomers of amino alcohols.

a, Different enantiomers give distinct activities in biological systems. **b**, Stereoisomeric amino alcohols (esters) and their biological activities. **c**, Bottom, our proposed one-pot synthesis protocol for forming all stereoisomers of amino alcohols via selective hydrosilylation/hydroamination reactions. The start point is an (*E*)- or (*Z*)-enal or enone, which is chemoselectively reduced by a CuH-based catalyst to yield corresponding allylic alcohols, which then undergo regio- and stereoselective hydroamination to afford amino-alcohol products bearing multiple contiguous stereocentres. Further details are in Fig. 2. Top, undesired potential side reactions. ‘L’ denotes ‘ligand’. Red bonds highlight different stereocentres.

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

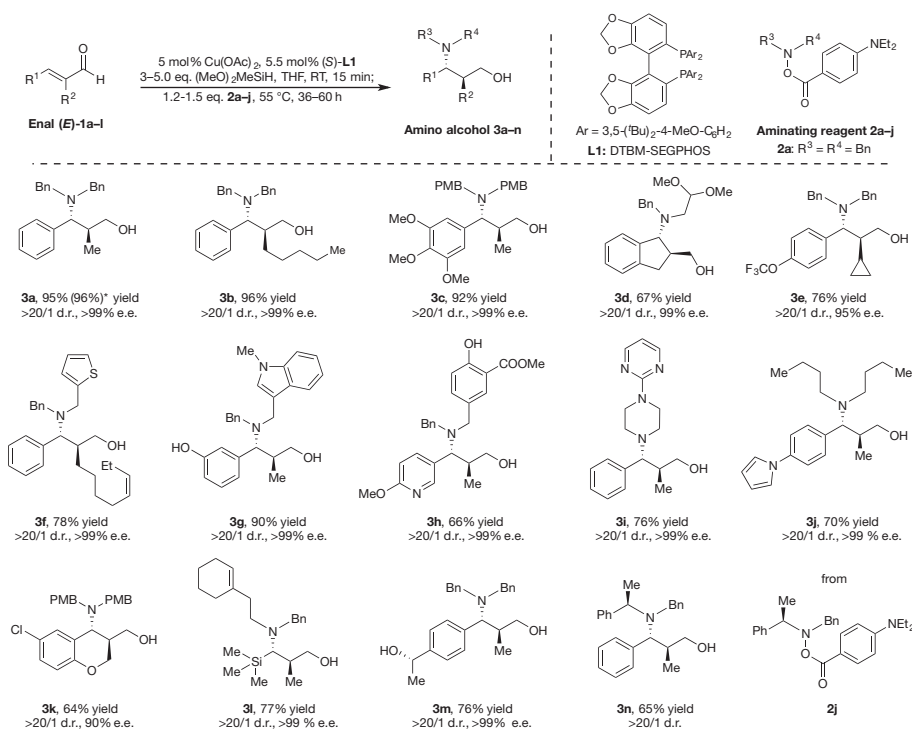


Figure 2 | Asymmetric hydrosilylation/hydroamination of enals.

Top row, the reaction studied here; bottom rows, substrate scope. Isolated yields are reported (average of two runs on 0.5 mmol scale). Diastereomeric ratios (d.r.) were determined by gas-chromatography (GC) and nuclear magnetic resonance (NMR) analysis. Enantiomeric excesses

(e.e.) were determined by high-performance liquid-chromatography (HPLC) analysis using chiral stationary phases. *Reaction was conducted on 5 mmol scale. See Supplementary Information for details. Bn, benzyl; PMB, *p*-methoxybenzyl; RT, room temperature; THF, tetrahydrofuran.

and electrophilic aminating reagents²⁷ could be applied to the synthesis of this class of compounds.

We hypothesized that enals and enones, which are readily available as geometrically pure isomers, could be ideal precursors to amino alcohols. In particular, we anticipated that, in one synthetic operation, a CuH-based catalyst could reduce enals and enones chemoselectively to the corresponding allylic alcohols, which would then undergo regio- and stereoselective hydroamination to afford amino-alcohol products bearing multiple contiguous stereocentres (Fig. 1c). Because of the synfacial nature of the hydroamination process, we reasoned that this two-step sequence would be stereospecific with respect to olefin geometry. Thus, effective catalyst control would allow all diastereomeric possibilities to be generated through the appropriate choice of substrate geometry (*E* or *Z*) and ligand enantiomer (*R* or *S*). Although the asymmetric hydrosilylation of ketones using a copper catalyst is a well known process²⁸, we were aware that 1,2-reduction of α,β -unsaturated carbonyl compounds by CuH is less favourable than 1,4-reduction, owing to the inherent preference of copper to coordinate to the olefin via soft-soft interactions²⁶. Previous work^{29,30} suggested that control of the regioselectivity in the CuH reduction of Michael acceptors was sensitive to subtle variations in the steric and electronic properties of the ligand and of the substituents of substrates. On the other hand, controlling the regio- and stereoselectivity of the hydroamination step is also non-trivial²⁵, because of the steric and electronic bias of the allylic silyl ether intermediates and potential matched/mismatched effects between substrates and chiral catalysts. The copper-based catalyst system described here alleviates these problems, providing access to all possible amino-alcohol stereoisomers in high chemo-, regio-, diastereo- and enantioselectivity, starting from readily available enals and enones.

In initial efforts to implement this proposed hydrosilylation/hydroamination sequence, we first treated (*E*)-2-methyl-cinnamaldehyde (Fig. 2; compound 1a) with an excess of dimethoxymethylsilane in the presence of 5 mol% of copper acetate and (*S*)-DTBM-SEGPHOS ((*S*)-(+)-5,5'-bis[di(3,5-di-*tert*-butyl-4-methoxyphenyl)]

phosphino]-4,4'-bi-1,3-benzodioxole; (*S*)-L1) at room temperature for 15 minutes. We observed neither the conventional 1,4-reduction product nor the over-reduced saturated alcohol, and the desired 1,2-adduct was obtained in nearly quantitative yield. Further treatment of the reaction mixture with aminating reagent 2a at 55 °C effectively provided the amino alcohol 3a with a 95% yield and with complete diastereo- and enantioselectivity (diastereomeric ratio (d.r.) >20/1; enantiomeric excess (e.e.) >99%). A thorough evaluation of the electrophilic amine source indicated that 2a was the optimal aminating reagent. We attributed the high efficiency of 2a to the presence of its electron-rich *para*-diethylaminobenzoyl group. This group resulted in slower reductive decomposition of 2a, and presumably faster regeneration of the CuH catalyst through σ -bond metathesis between the corresponding copper benzoate species and hydrosilane species^{22,23}. L1 was the best ligand among the phosphines screened in terms of both reactivity and selectivity (see Supplementary Information for details).

With these optimized reaction conditions, we investigated the substrate scope of this one-pot transformation. We found that an array of β -aryl-substituted (*E*)-enals could be efficiently transformed to the corresponding chiral amino alcohols in a highly regio-, diastereo- and enantioselective manner (Fig. 2, 3a–3n). A diverse range of hydroxylamine esters and enals with a variety of functional groups were suitable coupling partners for this sequential transformation, including phenols (3g, 3h), an aryl chloride (3k), an ester (3h), an acetal (3d), a trifluoromethoxyl (3e), *cis*-olefins (3f, 3l) and a trimethylsilyl group (from the reaction of a vinyl silane) (3i). In addition to acyclic substrates, a cyclic enal and cyclic aminating reagents were both compatible, providing the products with high levels of stereocontrol (3d, 3i, 3k). Furthermore, substrates bearing a broad range of pharmaceutically important heteroaromatic components—including an indole (3g), a thiophene (3f), a pyridine (3h), a pyrrole (3j), a pyrimidylpiperidine (3i), and a chromene (3k)—could be readily converted to the desired products with excellent enantioselectivity. Additionally, an enal substrate bearing a ketone functional group smoothly underwent

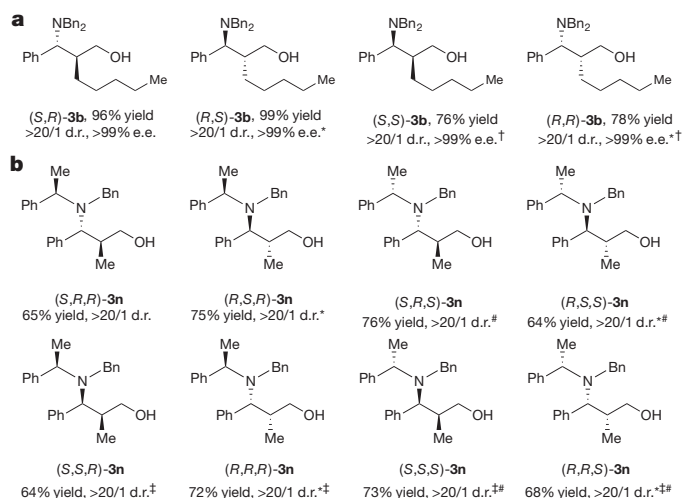


Figure 3 | All stereoisomers of amino alcohols from enals. **a**, Reaction using (*S*)-**L1** and (*E*)-2-amyl-cinnamaldehyde (**1b**) under the standard conditions. Asterisks denote reactions using (*R*)-**L1** instead of (*S*)-**L1**. The dagger denotes a reaction using (*Z*)-**1b** instead of (*E*)-**1b**. **b**, Reaction using (*S*)-**L1**, (*E*)-2-methyl-cinnamaldehyde (**1a**) and **2j** under the standard conditions. Asterisk denotes the use of (*R*)-**L1** instead of (*S*)-**L1**. The double dagger denotes the use of (*Z*)-**1a** instead of (*E*)-**1a**. The hash symbol indicates reactions using *ent*-**2j** instead of **2j**.

the double hydrosilylation and hydroamination sequence under these conditions, furnishing amino diol (**3m**) bearing three stereocentres as a single stereoisomer. Lastly, a reaction conducted on a 5 mmol scale with decreased catalyst loading (1 mol%) efficiently provided the desired product (**3a**) in undiminished yield and stereoselectivity, demonstrating the robustness and practicality of this process.

As described above, we were particularly interested in applying this approach to access all of the possible stereoisomers selectively, with high diastereo- and enantiocontrol for a given substrate. We felt that using the (*Z*)-enal substrate would produce the pair of diastereomers that is complementary to the pair realized from the (*E*)-enal substrate. Accordingly, we subjected (*Z*)-2-amyl-cinnamaldehyde (**1b**) to the reaction conditions described above, and produced the desired diastereomeric *syn*-amino alcohol in high chemical yield with complete diastereo- and enantioselectivity. Thus, by correctly combining the appropriate enantiomer of the CuH catalyst's ligand with the olefin geometry of the enal substrate, we could readily prepare all four stereoisomers of the corresponding amino alcohol with full control of absolute and relative stereochemistry (Fig. 3a).

We also found that amino alcohols bearing three stereogenic centres could be generated with excellent catalyst-controlled diastereoselectivity when we used chiral hydroxylamine esters. The existing α -stereocentre on chiral aminating reagents did not interfere with the stereoselectivity when applied to our catalyst system. Thus, all eight stereoisomers of the amino alcohol **3n** were easily constructed with good yield and excellent diastereoselectivity in one step, by selecting the appropriate enantiomer of the chiral aminating reagent and geometric isomer of the enal substrate as starting materials and using either enantiomer of the chiral catalyst (Fig. 3b).

We then sought to examine the possibility of using enones as substrates for the rapid synthesis of chiral amino alcohols with three contiguous stereocentres. Basing our protocol on Lipshutz and colleagues' work³⁰ on the asymmetric CuH-catalysed 1,2-reduction of enones, we found that the readily accessible enone **4a** could be effectively converted to the corresponding chiral allylic alcohol in quantitative yield and 92% e.e. at -60°C in the presence of 5 mol% of $\text{Cu}(\text{OAc})_2$ -**L1** complex. We added the aminating reagent to the reaction mixture while heating the mixture at 55°C , furnishing compound **5a** at 76% chemical yield with complete diastereo- and enantioselectivity (>20/1 d.r.; >99% e.e.). Under these reaction conditions, a variety of enones were transformed

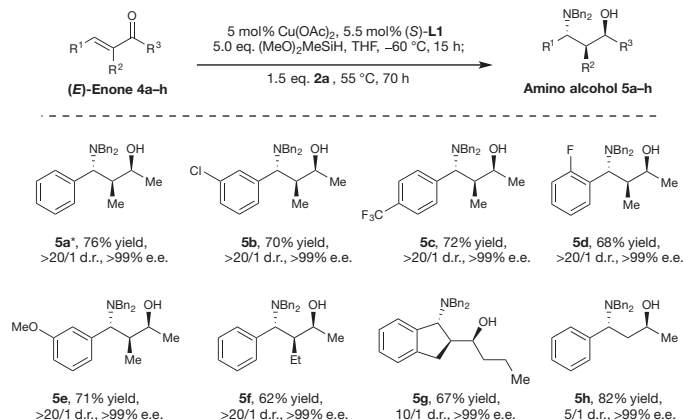


Figure 4 | Asymmetric hydrosilylation/hydroamination of enones.

Top row, the reaction studied here; bottom rows, substrate scope. Isolated yields are reported (average of two runs on 1.0 mmol scale). Diastereomeric ratios (d.r.) were determined by GC and NMR analysis. Enantiomeric excesses (e.e.) were determined by HPLC analysis. The asterisk indicates where the absolute and relative stereochemistry of **5a** was determined to be (*S,S,R*) by single-crystal X-ray diffraction. See Supplementary Information for details.

successfully to the respective amino-alcohol products (Fig. 4, **5a–5f**) with excellent absolute and relative stereoselectivity (>20/1 d.r., >99% e.e.). In addition, a cyclic enone was converted into indanyl amino alcohol (**5g**) with high diastereoselectivity (10/1 d.r.) and outstanding enantioselectivity (>99% e.e.) for both diastereomers. Lastly, we found that α -substitution on the enone was not crucial for selective 1,2-reduction. For instance, the less-substituted product **5h** could be obtained with high enantioselectivity and a synthetically useful diastereomeric ratio from benzylidenacetone.

Subsequently, we wondered whether these conditions could be adapted to the stereodivergent construction of all eight stereoisomers of **5a**. Above (Fig. 4), we prepared (*S,S,R*)-**5a** from (*E*)-**4a** in a one-pot sequence by using (*S*)-**L1** as the ligand. To prepare (*S,R,S*)-**5a**, we developed a modified protocol involving a ligand switch in which (*E*)-**4a** was first hydrosilylated using (*S*)-**L1** as the ligand, and then the resulting chiral allylic alcohol was isolated and subjected to hydroamination conditions using a copper catalyst based on (*R*)-**L1**. Use of enantiomeric ligands for the one-pot and ligand-switch protocols yielded (*R,R,S*)-**5a** and (*R,S,R*)-**5a**, respectively, allowing four of the possible stereoisomers of **5a** to be prepared from (*E*)-**4a**. We sought to prepare the four remaining stereoisomers of **5a** from (*Z*)-**4a** by using a similar strategy. The same ligand-switch protocol applied to (*Z*)-**4a** furnished (*R,S,S*)-**5a** and (*S,R,R*)-**5a**. Initial attempts to prepare (*S,S,S*)-**5a** and (*R,R,R*)-**5a** were unsuccessful, presumably owing to the unfavourable steric interactions between the intermediate chiral (*Z*)-allylic alcohol and the **L1**-based catalyst. Accordingly, we selected a less bulky ligand, DM-SEGPHOS (5,5'-bis[di(3,5-xylyl)phosphino]-4,4'-bi-1,3-benzodioxole; **L2**), for the hydroamination step to obtain these products. Ultimately, all eight stereoisomers of **5a** were prepared expediently in one to two steps with useful isolated yields (33% to 76%), with complete enantioselectivity (>99% e.e.) and good to excellent diastereoselectivity (7/1 to >20/1 produced in the reaction mixture before purification) (Fig. 5a). Upon isolation and chromatography, all isomers were obtained at >20/1 d.r., as confirmed by the high-performance liquid-chromatography traces in Fig. 5b. Thus, the enantioselective hydroamination steps proceeded through excellent catalyst control in all eight cases.

We have developed a unified and stereodivergent strategy for rapidly and predictably constructing amino alcohols that allows all possible stereoisomers of a given product to be synthesized. This protocol assembles two or three contiguous stereocentres, using enal and enone substrates, in a highly selective, copper-catalysed hydrosilylation/hydroamination sequence. Essential to our approach were highly effective catalyst control and the application of a stereospecific process

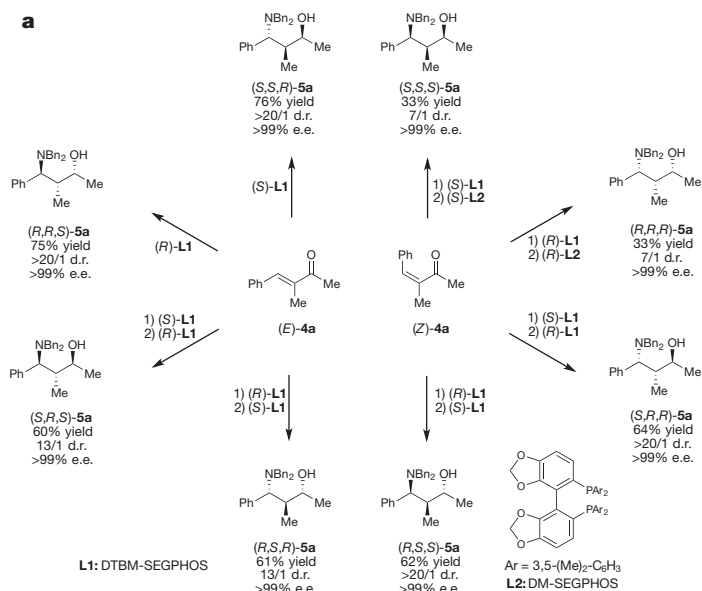


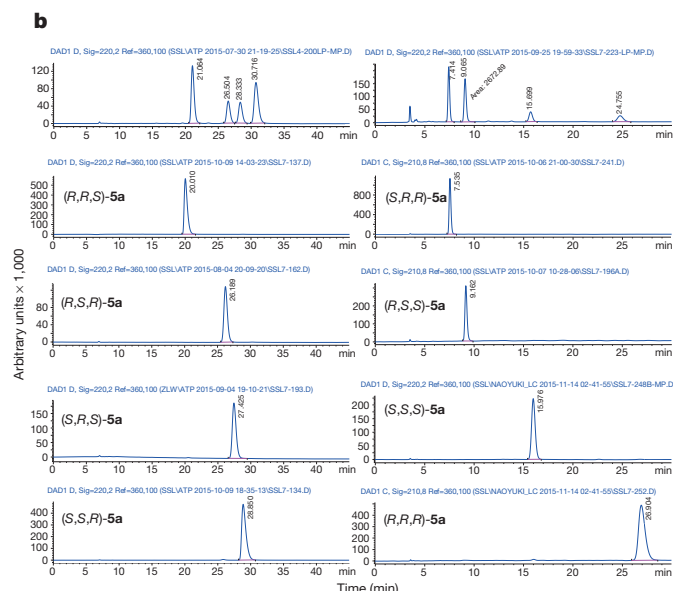
Figure 5 | All eight stereoisomers of amino alcohols synthesized from enones. **a.** Access to all stereoisomers via a reaction using (*E*)- or (*Z*)-**4a**, showing the catalyst permutations in each step. Isolated yields are

to readily obtained pure geometric alkene isomers—factors that might be generally applicable to the development of other enantio- and diastereodivergent hydrofunctionalization reactions.

Received 24 November 2015; accepted 27 January 2016.

Published online 28 March 2016.

1. Jozwiak, K., Lough, W. J. & Wainer, I. W. *Drug Stereochemistry: Analytical Methods and Pharmacology* 3rd edn (Informa, 2012).
2. Wermuth, C. G. *The Practice of Medicinal Chemistry* 3rd edn (Elsevier, 2008).
3. Jacobsen, E. N., Pfaltz, A. & Yamamoto, H. (eds) *Comprehensive Asymmetric Catalysis* Vol. I–III, Suppl. I–II (Springer, 1999).
4. Tian, X. *et al.* Diastereodivergent asymmetric sulfa-Michael additions of α -branched enones using a single chiral organic catalyst. *J. Am. Chem. Soc.* **133**, 17934–17941 (2011).
5. McInturff, E. L., Yamaguchi, E. & Krische, M. J. Chiral-anion-dependent inversion of diastereo- and enantioselectivity in carbonyl crotylation via ruthenium-catalyzed butadiene hydrohydroxyalkylation. *J. Am. Chem. Soc.* **134**, 20628–20631 (2012).
6. Wang, B., Wu, F., Wang, Y., Liu, X. & Deng, L. Control of diastereoselectivity in tandem asymmetric reactions generating nonadjacent stereocenters with bifunctional catalysis by cinchona alkaloids. *J. Am. Chem. Soc.* **129**, 768–769 (2007).
7. Yan, X.-X. *et al.* Highly diastereoselective switchable enantioselective Mannich reaction of glycine derivatives with imines. *J. Am. Chem. Soc.* **130**, 14362–14363 (2008).
8. Luparia, M. *et al.* Catalytic asymmetric diastereodivergent deracemization. *Angew. Chem. Int. Edn* **50**, 12631–12635 (2011).
9. Lu, G., Yoshino, T., Morimoto, H., Matsunaga, S. & Shibasaki, M. Stereodivergent direct catalytic asymmetric mannich-type reactions of α -isothiocyanato ester with ketimines. *Angew. Chem. Int. Edn* **50**, 4382–4385 (2011).
10. Mechler, M. & Peters, R. Diastereodivergent asymmetric 1,4-addition of oxindoles to nitroolefins by using polyfunctional nickel-hydrogen-bond-azoliun catalysts. *Angew. Chem. Int. Edn* **54**, 10303–10307 (2015).
11. Lee, E. C., Hodous, B. L., Bergin, E., Shih, C. & Fu, G. C. Catalytic asymmetric Staudinger reactions to form β -lactams: an unanticipated dependence of diastereoselectivity on the choice of the nitrogen substituent. *J. Am. Chem. Soc.* **127**, 11586–11587 (2005).
12. Huber, J. D. & Leighton, J. L. Highly enantioselective imine cinnamylation with a remarkable diastereoselective switch. *J. Am. Chem. Soc.* **129**, 14552–14553 (2007).
13. Huang, Y., Walji, A. M., Larsen, C. H. & MacMillan, D. W. C. Enantioselective organo-cascade catalysis. *J. Am. Chem. Soc.* **127**, 15051–15053 (2005).
14. Simmons, B., Walji, A. M. & MacMillan, D. W. C. Cycle-specific organocascade catalysis: application to olefin hydroamination, hydro-oxidation, and amino-oxidation, and to natural product synthesis. *Angew. Chem. Int. Edn* **48**, 4349–4353 (2009).
15. Krautwald, S., Sarlah, D., Schafroth, M. A. & Carreira, E. M. Enantio- and diastereodivergent dual catalysis: α -allylation of branched aldehydes. *Science* **340**, 1065–1068 (2013).
16. Schindler, C. S. & Jacobsen, E. N. A new twist on cooperative catalysis. *Science* **340**, 1052–1053 (2013).
17. Nugent, T. C. *Chiral Amine Synthesis: Methods, Developments and Applications* (Wiley VCH, 2010).
18. Zhu, S., Niljanskul, N. & Buchwald, S. L. Enantio- and regioselective CuH-catalyzed hydroamination of alkenes. *J. Am. Chem. Soc.* **135**, 15746–15749 (2013).



reported. Diastereomeric ratios (d.r.) were determined by NMR analysis using a crude reaction mixture. **b**, HPLC traces of all stereoisomeric amino-alcohol samples. See Supplementary Information for details.

19. Shi, S.-L. & Buchwald, S. L. Copper-catalyzed selective hydroamination reactions of alkynes. *Nature Chem.* **7**, 38–44 (2015).
20. Zhu, S. & Buchwald, S. L. Enantioselective CuH-catalyzed anti-Markovnikov hydroamination of 1, 1-disubstituted alkenes. *J. Am. Chem. Soc.* **136**, 15913–15916 (2014).
21. Yang, Y., Shi, S.-L., Niu, D., Liu, P. & Buchwald, S. L. Catalytic asymmetric hydroamination of unactivated internal olefins to aliphatic amines. *Science* **349**, 62–66 (2015).
22. Niu, D. & Buchwald, S. L. Design of modified amine transfer reagents allows the synthesis of α -chiral secondary amines via CuH-catalyzed hydroamination. *J. Am. Chem. Soc.* **137**, 9716–9721 (2015).
23. Bandar, J. S., Pirnot, M. T. & Buchwald, S. L. Mechanistic studies lead to dramatically improved reaction conditions for the Cu-catalyzed asymmetric hydroamination of olefins. *J. Am. Chem. Soc.* **137**, 14812–14818 (2015).
24. Miki, Y., Hirano, K., Satoh, T. & Miura, M. Copper-catalyzed intermolecular regioselective hydroamination of styrenes with polymethylhydrosiloxane and hydroxylamines. *Angew. Chem. Int. Edn* **52**, 10830–10834 (2013).
25. Hannedouche, J. & Schulz, E. Asymmetric hydroamination: a survey of the most recent developments. *Chemistry* **19**, 4972–4985 (2013).
26. Deutsch, C., Krause, N. & Lipshutz, B. H. CuH-catalyzed reactions. *Chem. Rev.* **108**, 2916–2927 (2008).
27. Barker, T. J. & Jarvo, E. R. Developments in transition-metal-catalyzed reactions using electrophilic nitrogen sources. *Synthesis* **24**, 3954–3964 (2011).
28. Lipshutz, B. H., Noson, K., Chrisman, W. & Lower, A. Asymmetric hydrosilylation of aryl ketones catalyzed by copper hydride complexed by nonracemic biphenyl bis-phosphine ligands. *J. Am. Chem. Soc.* **125**, 8779–8789 (2003).
29. Chen, J.-X. et al. Highly chemoselective catalytic hydrogenation of unsaturated ketones and aldehydes to unsaturated alcohols using phosphine-stabilized copper(I) hydride complexes. *Tetrahedron* **56**, 2153–2166 (2000).
30. Moser, R., Boskovic, Z. V., Crowe, C. S. & Lipshutz, B. H. CuH-catalyzed enantioselective 1,2-reductions of α,β -unsaturated ketones. *J. Am. Chem. Soc.* **132**, 7852–7853 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the National Institutes of Health (grant GM-58160 to S.L.B.). The content of this paper is solely our responsibility and does not necessarily represent the official views of the National Institutes of Health. We thank Y.-M. Wang and M. T. Pirnot for help preparing the manuscript. The departmental X-ray diffraction instrumentation was purchased with the help of funding from the National Science Foundation (CHE-0946721). We thank P. Müller for X-ray crystallographic analysis of compound **5a**.

Author Contributions S.-L.S. and S.L.B. conceived the idea and designed the research. S.-L.S. and Z.L.W. performed the experiments. S.-L.S. and S.L.B. wrote the manuscript. All authors commented on the final draft of the manuscript and contributed to the analysis and interpretation of the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L.B. (sbuchwal@mit.edu).

Recent improvement and projected worsening of weather in the United States

Patrick J. Egan^{1*} & Megan Mullin^{2*}

As climate change unfolds, weather systems in the United States have been shifting in patterns that vary across regions and seasons^{1–7}. Climate science research typically assesses these changes by examining individual weather indicators, such as temperature or precipitation, in isolation, and averaging their values across the spatial surface. As a result, little is known about population exposure to changes in weather and how people experience and evaluate these changes considered together. Here we show that in the United States from 1974 to 2013, the weather conditions experienced by the vast majority of the population improved. Using previous research on how weather affects local population growth^{8–14} to develop an index of people's weather preferences, we find that 80% of Americans live in counties that are experiencing more pleasant weather than they did four decades ago. Virtually all Americans are now experiencing the much milder winters that they typically prefer, and these mild winters have not been offset by markedly more uncomfortable summers or other negative changes. Climate change models predict that this trend is temporary, however, because US summers will eventually warm more than winters. Under a scenario in which greenhouse gas emissions proceed at an unabated rate (Representative Concentration Pathway 8.5), we estimate that 88% of the US public will experience weather at the end of the century that is less preferable than weather in the recent past. Our results have implications for the public's understanding of the climate change problem, which is shaped in part by experiences with local weather^{15–20}. Whereas weather patterns in recent decades have served as a poor source of motivation for Americans to demand a policy response to climate change, public concern may rise once people's everyday experiences of climate change effects start to become less pleasant.

A growing body of work, synthesized most recently in the 2014 National Climate Assessment, documents how global climate change is altering US weather systems¹. Average surface temperatures have increased, and recent decades have seen a rise in the frequency of extreme heat events and a decrease in cold spells^{2–5}. Patterns in these trends vary across regions within the United States as well as across seasons. Thus far, temperatures have risen more in winters than in summers^{6,7}, producing in most regions a climate that is more temperate than it had been previously. On the basis of findings in the economics literature indicating that Americans prefer to live in temperate climates^{8–14}, we hypothesized that Americans have experienced an improvement in their local weather over recent decades, which may help explain low levels of concern about the problem among the mass public.

Work assessing the impact of climate change on weather typically uses spatially averaged changes in individual weather indicators, such as temperature or precipitation. Our approach is to examine change in the weather on the basis of population-weighted exposure and to evaluate this change in light of the trade-offs people have been shown to make among different weather conditions. Our focus on weather

preferences requires analyses that account for seasonal variation in warming patterns. Moreover, recognizing that Americans have limited understanding of the likely effects of climate change, our analysis includes shifts in weather indicators demonstrated to be meaningful to the mass public in addition to those theorized by climate scientists to be most associated with global warming.

We investigated our hypothesis using 40 years of daily weather data reported in the Global Surface Summary of the Day (GSOD) and the US Historical Climatology Network (USHCN) data sets^{21,22} (Supplementary Table 1 and Extended Data Fig. 1). After assigning weather stations to counties based on station distance from county population centroids, we calculated annual means of weather indicators weighted by 2010 county population (Extended Data Table 1). We used constant population weights—rather than adjusting them for shifts over time in county population—to isolate the impact of weather trends from any changes in aggregate exposure attributable to population growth or migration²³.

As objective measures of the average American's weather preferences, we used estimates of revealed preferences produced by a sizeable literature analysing the effects of local climate on population growth^{8–14}. These studies estimate change over time in population or net migration at the city, county or metropolitan-area level as a function of normal weather and a set of control variables. We used each study's coefficients on weather-related variables to assign regression scores to our county-level weather data. With these values, we calculated an annual weather preference index (WPI) score for each county, an indicator of the average American's revealed preference for different types of weather conditions. The WPI's unit of measure is the *ceteris paribus* expected rate of population change associated with that year's weather. In using these studies for our purposes, we assume that, all things being equal, Americans move to, and continue to reside in, places with local climates that they prefer. We further assume that aggregate tastes for weather were stable over the period between 1974 and 2013, and that controls in published models of weather-related population growth remove any confounding between weather tastes and the intensity of those tastes or ease of residential mobility. Our focus is only on preferences for long-term weather conditions, not on less predictable public responses to extreme weather events.

We report here on the WPI developed from estimates in a widely cited study that includes the most comprehensive models published in this literature¹⁴. Results were consistent using WPIs derived from parameter estimates in every other published study of this type^{9,11–13} (Extended Data Tables 2 and 3). In the model used here, county population growth from 1970 to 2000 is a function of five long-term normal weather indicators: January average daily maximum temperature; July daily heat index (derived from July average daily maximum temperature and July average daily mean relative humidity); July average daily mean relative humidity (entered into the model separately); annual precipitation; and the number of days on which precipitation occurs annually.

¹Wilf Family Department of Politics, New York University, New York, New York 10012, USA. ²Nicholas School of the Environment, Duke University, Durham, North Carolina 27708, USA.

*These authors contributed equally to this work.

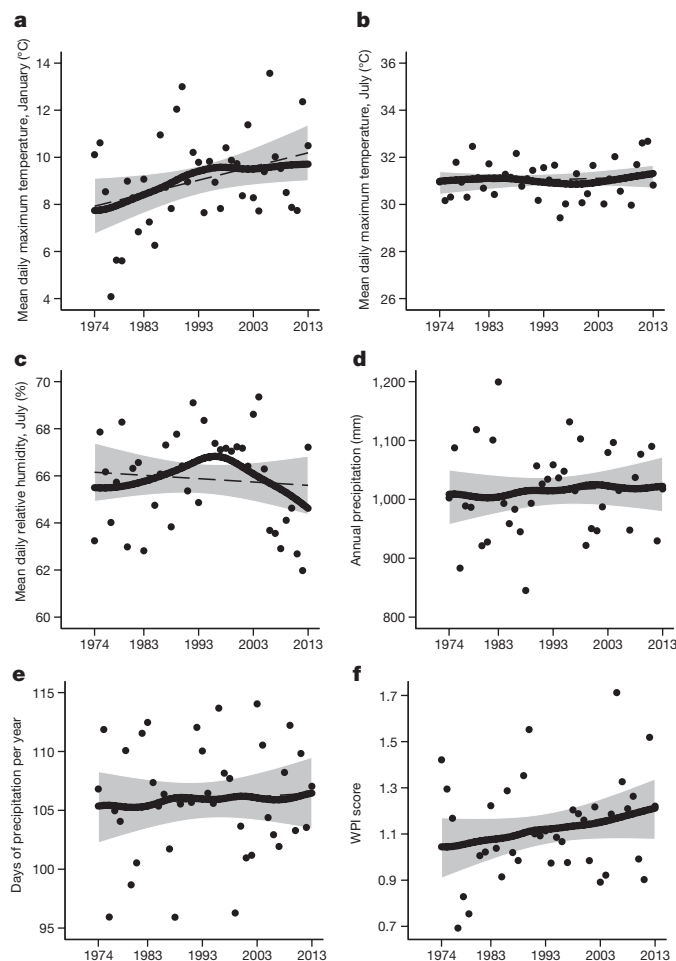


Figure 1 | Weather as experienced by the US population, 1974–2013.

a–f, Annual means of county-level January mean daily maximum temperature (**a**), July mean daily maximum temperature (**b**), July mean daily relative humidity (**c**), annual precipitation amount (**d**), annual precipitation days (**e**) and WPI score (**f**), weighted by 2010 county population. Solid lines are locally weighted scatterplot smoothing (LOWESS) smoothers at 0.8 bandwidth; dashed lines display linear fits with shaded 95% confidence intervals.

Calculations of annual trends for these five indicators (Fig. 1) confirmed that Americans experienced pronounced winter warming between 1974 and 2013: a regression of daily maximum January temperatures on year estimated a population-weighted average increase of 0.58 °C per decade. By contrast, summer warming trends were much less substantial: daily maximum July temperatures rose by only 0.07 °C per decade. Summer relative humidity rose in the mid-1990s

but since then has declined in an abrupt pattern noted elsewhere²⁴. As a result, summer heat-index values—calculated from temperatures and humidity—have remained relatively stable, with July heat index values rising only by the equivalent of 0.07 °C per decade. Annual precipitation rose slightly (increasing by 5.64 mm per decade) and occurred a bit more frequently (0.31 additional days per decade). In a complementary analysis, we estimated within-county change over the four-decade period with county-by-county regressions of each weather indicator on year (Table 1). These estimates reflected aggregate trends: since 1974, winter temperatures have become warmer for virtually all Americans; within-county change in the other indicators has been much less consistent.

We measured Americans' preferences towards the weather they have been experiencing by using coefficient estimates from the population growth models to calculate WPI scores for each county in each year between 1974 and 2013. Details on our calculations appear in Methods. All published population growth models reveal the same preference structure with respect to seasonal temperatures: Americans appreciate warm winters and dislike hot, humid summers. The most comprehensive WPI model indicates that these preferences become more pronounced at extreme weather values¹⁴. Preferences about precipitation are mixed and have modest substantive impact, but Americans typically prefer less precipitation spread out over more days. Face validity of the WPI measure was confirmed by the geographical distribution of county mean scores (Fig. 2a).

Analyses of the changes in these scores demonstrate that over the past 40 years, US weather has become more preferable for most Americans. The trend in the population-weighted annual means of county WPI scores (Fig. 1f) was in a steadily upward direction over the entire period at a population growth rate equivalent of 0.04 per decade. Estimated trends derived from county-by-county regressions of WPI on year indicated that improved weather is prevalent across most of the country (Fig. 2b), including areas covered by more weather stations, for which measurement is less susceptible to error (Extended Data Fig. 2). Currently, 80% of Americans live in counties where weather has improved over the past four decades (Table 1). Sensitivity analyses using a variety of measurement choices and WPI models produced estimates of the population exposed to improved weather ranging from 72% to 96% (Extended Data Tables 2–5).

However, climate models project that these seasonal trends will eventually reverse, and that future US warming will be more severe in summer than in winter²⁵. To estimate Americans' preferences regarding these weather changes, we used projections from models in phase 5 of the Coupled Model Intercomparison Project (CMIP5) that have been downscaled and aggregated to the county level in the US Geological Survey's National Climate Change Viewer^{26,27}. We calculated the projected change in WPI scores from the observed 1974–2013 mean values to the year 2099 under two Representative Concentration Pathway (RCP) scenarios, fixing all non-temperature weather indicators at their means for the final 10 years of our study

Table 1 | Change in county weather indicators per decade, 1974–2013

Indicator	25th percentile	50th percentile	75th percentile	Un-weighted mean	Mean weighted by 1970 population	Mean weighted by 2010 population	Percentage experiencing increase, weighted by 2010 population
Maximum daily January temperature (°C)	0.55	0.70	0.85	0.69	0.62	0.58	99.9%
Maximum daily July temperature (°C)	−0.14	0.02	0.15	0.03	0.06	0.07	60.1%
July relative humidity (%)	−0.61	−0.12	0.49	0.01	−0.12	−0.14	38.5%
Maximum daily July heat index (°C equivalent)	−0.23	0.04	0.31	0.03	0.07	0.07	57.4%
Annual precipitation (mm)	−12.21	7.83	22.64	5.36	10.63	5.64	60.1%
Annual precipitation (days)	−0.99	0.38	1.70	0.31	0.72	0.31	63.8%
WPI score (annual population growth rate equivalent)	0.00	0.05	0.09	0.05	0.04	0.04	79.6%

Shown are summary statistics of within-county regression trend estimates for 3,037 counties in the contiguous United States using temperature and humidity data reported by 324 GSOD weather stations and precipitation data from 601 USHCN weather stations.

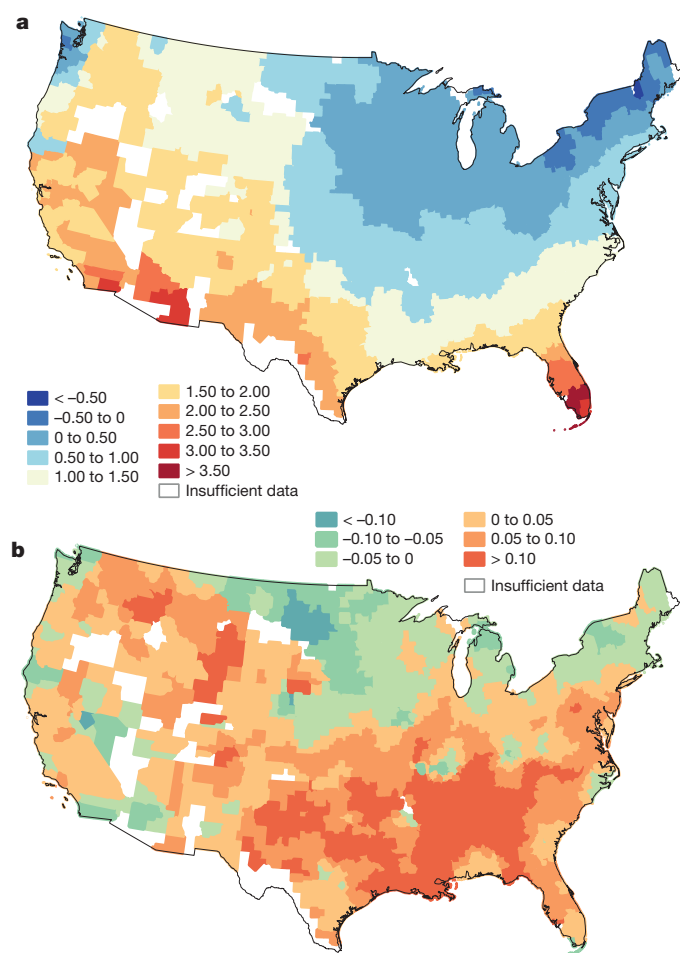


Figure 2 | WPI score by county, 1974–2013. a, Mean WPI, derived using calculations from equation (1) in Methods with 40-year mean values of weather indicators. **b,** Population growth rate equivalent change in WPI by decade, derived from county-by-county regressions of annual WPI on year as summarized in Table 1.

period. In the first scenario, greenhouse gas emissions diminish to the extent that total radiative forcing is stabilized shortly after the year 2100 (RCP4.5). In the second, emissions instead proceed at an unabated rate (RCP8.5)²⁸. Under both scenarios, mean county WPI scores weighted by 2010 county population are projected to decline relative to the observed historical trend—and dramatically so under RCP8.5 (Fig. 3a), in which we estimate that 88% of the US public will experience less pleasant weather at the end of the century than in the recent past (Extended Data Table 6). The projected decline in weather pleasantness persists after adjusting estimated future WPI upward to reflect discrepancies between the observed temperature record and climate model simulations of historical temperatures⁷ (Fig. 3b).

Our analysis is necessarily limited to evaluating population exposure to weather conditions in the United States, where ample data availability has made possible the research on long-term weather and population mobility that provides our measure of the average American's preferences about weather. Although we cannot assume that the same preference structure holds outside of the United States, the temperate zone of the Northern Hemisphere is where people are most likely to share a similar taste for weather. Surface-averaged projections from CMIP5 models indicate that seasonal changes will vary in this zone²⁹. Europe is similar to the United States in that summers are expected to warm as much or more than winters up to the year 2100; by contrast, in Canada and Russia—and to a lesser extent, China—weather is projected to become more temperate (Extended Data Table 7).

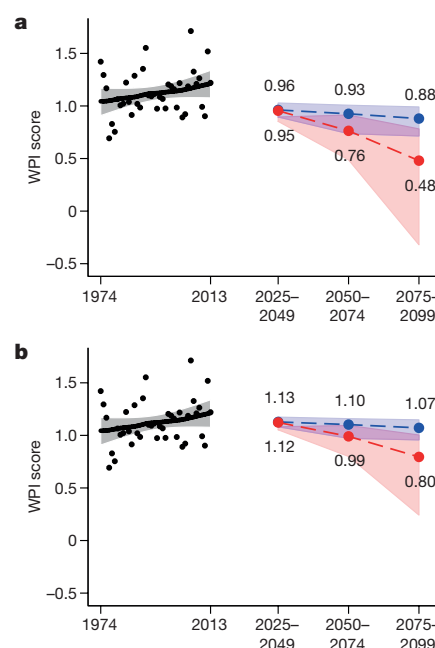


Figure 3 | Historical and projected trends in pleasantness of US weather. a, b, Black line is the smoothed historical trend (and grey area a 95% confidence interval around a linear fit) in WPI scores experienced by the US population, 1974–2013. Coloured areas depict projected scores under RCP4.5 (blue) and RCP8.5 (red). Dashed lines show the median forecast from 30 CMIP5 models, and shaded areas indicate the range from the 10th to 90th percentile values. **a,** Projections calculated directly from CMIP5 model output. **b,** Projections adjusted to reflect discrepancies between observed historical data and modelled output for the historical period. Analysis includes the contiguous US counties.

Our results have implications for how the complex phenomenon of climate change should be explained to the public. The scientific community's emphasis on mean temperatures averaged over space and across seasons does not correspond with how people actually experience the weather³⁰, which varies geographically and is interpreted through preference structures that value warmer temperatures differently at different times of year. In the United States, people's experiences with daily weather since the time that they first heard about climate change have thus far mostly been positive. For the large portion of the public who form their beliefs and concern about climate change in part on the basis of their direct experience with weather^{15–20}, the changes to which they have been exposed to date cannot be relied upon to provide the motivation needed to overcome apathy in responding to global warming. Our projections indicate that this may change in the decades to come as trends shift and weather becomes less pleasant. Both now and in the future, the public's evaluations of the weather changes brought about by climate change should be taken into account by scientists and others seeking to explain the problem, mobilize concern and catalyse policy response.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 March 2015; accepted 16 February 2016.

- Melillo, J. M., Richmond, T. C. & Yohe, G. W. *Climate Change Impacts in the United States: The Third National Climate Assessment* (US Global Change Research Program, 2014).
- Gaffen, D. J. & Ross, R. J. Increased summertime heat stress in the US. *Nature* **396**, 529–530 (1998).
- Peterson, T. C. *et al.* Monitoring and understanding changes in heat waves, cold waves, floods, and droughts in the United States: state of knowledge. *Bull. Am. Meteorol. Soc.* **94**, 821–834 (2013).
- Menne, M. J., Williams, C. N. Jr & Vose, R. S. The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull. Am. Meteorol. Soc.* **90**, 993–1007 (2009).

5. Oswald, E. M. & Rood, R. B. A trend analysis of the 1930–2010 extreme heat events in the continental United States. *J. Appl. Meteorol. Climatol.* **53**, 565–582 (2014).
6. Kunkel, K. E. *et al.* *Regional Climate Trends and Scenarios for the U.S. National Climate Assessment: Part 9, Climate of the Contiguous United States* (National Oceanic and Atmospheric Administration, 2013).
7. Kumar, S., Kinter, J. III, Dirmeyer, P. A., Pan, Z. & Adams, J. Multidecadal climate variability and the “warming hole” in North America: results from CMIP5 twentieth- and twenty-first-century climate simulations. *J. Clim.* **26**, 3511–3527 (2013).
8. Glaeser, E. L., Kolko, J. & Saiz, A. Consumer city. *J. Econ. Geogr.* **1**, 27–50 (2001).
9. Glaeser, E. L. & Shapiro, J. M. Urban growth in the 1990s: is city living back? *J. Reg. Sci.* **43**, 139–165 (2003).
10. Glaeser, E. L. Reinventing Boston: 1640–2003. *J. Econ. Geogr.* **5**, 119–153 (2005).
11. Mueser, P. R. & Graves, P. E. Examining the role of economic opportunity and amenities in explaining population redistribution. *J. Urban Econ.* **37**, 176–200 (1995).
12. McGranahan, D. A. *Natural Amenities Drive Rural Population Change* (Economic Research Service, US Department of Agriculture, 1999).
13. Partridge, M. D. The dueling models: NEG vs amenity migration in explaining US engines of growth. *Pap. Reg. Sci.* **89**, 513–536 (2010).
14. Rappaport, J. Moving to nice weather. *Reg. Sci. Urban Econ.* **37**, 375–398 (2007).
15. Deryugina, T. How do people update? The effects of local weather fluctuations on beliefs about global warming. *Clim. Change* **118**, 397–416 (2013).
16. Egan, P. J. & Mullin, M. Turning personal experience into political attitudes: the effect of local weather on Americans’ perceptions about global warming. *J. Polit.* **74**, 796–809 (2012).
17. Howe, P. D., Markowitz, E. M., Lee, T. M., Ko, C.-Y. & Leiserowitz, A. Global perceptions of local temperature change. *Nature Clim. Chang.* **3**, 352–356 (2013).
18. Joireman, J. A., Truelove, H. B. & Duell, B. Effect of outdoor temperature, heat primes and anchoring on belief in global warming. *J. Environ. Psychol.* **30**, 358–367 (2010).
19. Li, Y., Johnson, E. J. & Zaval, L. Local warming: daily temperature change influences belief in global warming. *Psychol. Sci.* **22**, 454–459 (2011).
20. Zaval, L., Keenan, E. A., Johnson, E. J. & Weber, E. U. How warm days increase belief in global warming. *Nature Clim. Chang.* **4**, 143–147 (2014).
21. National Oceanic and Atmospheric Administration. *Global Surface Summary of the Day* (National Centers for Environmental Information, 2014).
22. Menne, M. J., Williams, C. N. Jr & Vose, R. S. *United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data* (Carbon Dioxide Information Analysis Center, 2015).
23. Cragg, M. I. & Kahn, M. E. Climate consumption and climate pricing from 1940 to 1990. *Reg. Sci. Urban Econ.* **29**, 519–539 (1999).
24. Brown, P. J. & DeGaetano, A. T. Trends in US surface humidity, 1930–2010. *J. Appl. Meteorol. Climatol.* **52**, 147–163 (2013).
25. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nature Clim. Chang.* **2**, 775–779 (2012).
26. Thrasher, B. *et al.* New downscaled climate projections suitable for resource management in the US. *Eos Trans. AGU* **94**, 321–323 (2013).
27. Alder, J. R. & Hostetler, S. W. *USGS National Climate Change Viewer* (US Geological Survey, 2013).
28. van Vuuren, D. P. *et al.* The representative concentration pathways: an overview. *Clim. Change* **109**, 5–31 (2011).
29. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
30. Lehner, F. & Stocker, T. F. From local perception to global perspective. *Nature Clim. Chang.* **5**, 731–734 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Alder, S. Anderson, A. Barreca, N. Beck, C. Dawes, S. Gordon, M. Hetherington, S. McDermid, J. Rappaport, M. Siegal and M. Smith.

Author Contributions P.J.E. and M.M. conceived the study. M.M. compiled the geographical data and weather data. P.J.E. performed the modelling. P.J.E. and M.M. analysed the results and wrote the paper.

Author Information The data and code needed to reproduce results have been deposited at P.J.E.’s Harvard University Dataverse website (<https://dataverse.harvard.edu/dataverse/patrickegan>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.J.E. (patrick.egan@nyu.edu).

METHODS

Data sources. Our historical weather data came from two sources: the Global Surface Summary of the Day (GSOD) data set, version 7 (<https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod>; accessed 22 July 2014)²¹, and the US Historical Climatology Network (USHCN), version 2.5 (<http://www.ncdc.noaa.gov/oa/climate/research/ushcn/>; accessed 22 September 2015)²², both maintained by the National Centers for Environmental Information. Block group-level and county-level US Census data, including geographical boundary data, came from the Minnesota Population Center's National Historical Geographic Information System, version 2.0 (<https://data2.nhgis.org/main>; accessed 30 July 2014)³¹. We obtained county-level monthly temperature projections from the National Climate Change Viewer (NCCV) (http://www.usgs.gov/climate_landuse/clu_rd/nccv.asp; accessed through direct communication with J. Alder on 20 July 2015 and 3 December 2015)^{26,27}, a US Geological Survey product that takes downscaled climate scenarios prepared by NASA (the National Aeronautics and Space Administration) and averages the 800-m gridded temperature data to the county level. Our international projections data came from the Royal Netherlands Meteorological Institute's Climate Change Atlas (http://climexp.knmi.nl/plot_atlas_form.py; accessed 11 September 2015)³².

Measurement approaches. We limited our analysis to data from weather stations in the contiguous United States that operated continuously between 1974 and 2013. This 40-year period is long enough to minimize sensitivity to natural variability in weather data, and it begins at a point in time when the number of weather stations included in standard data sets and the completeness of the data they reported both increased. The timespan covers the entire history of Americans' exposure to the idea of climate change, allowing us to track how weather has shifted during the time when the public might have perceived such shifts as attributable to climate change.

Temperature and humidity measurement. Daily weather data on temperature and humidity came from the GSOD data set, produced by the National Centers for Environmental Information from hourly weather station observations contained in the Integrated Surface Hourly data set²¹. Of the various land-based weather station data sets that offer daily summary data, GSOD is the only one that includes weather records necessary to measure a location's relative humidity, which the urban economics literature has shown to be an important climate amenity driving regional population growth. Temperature and humidity records in our data set came from the GSOD's daily station records of mean and maximum temperatures and mean dew point temperature (from which we calculated daily relative humidity and, in turn, daily heat index values). We included in the study only those GSOD stations reporting valid data on each of these weather indicators for at least 50% of the days in each of the 480 months of our study period, reducing the total number of stations in the analysis from 672 to $n = 324$ (Supplementary Table 1). Raising the threshold for valid data from 50% to 75% produced similar results (Extended Data Table 4a). Our final data set included a small number ($n = 34$) of stations that were relocated to nearby sites at some point between 1974 and 2013. In each case, the site location changed no more than 10 m in elevation and 0.1 decimal degree in latitude or longitude, and data reported by the relocated stations covered the entire study period with no more than a 15-day gap. Running our main analysis after omitting these relocated stations produced similar results (Extended Data Table 4b).

Precipitation measurement. Synoptic reporting of weather conditions in the GSOD data set introduces error in the calculation of daily precipitation indicators, so our main analyses employ daily precipitation records from weather stations in the USHCN, a designated subset of the National Oceanic and Atmospheric Association's Cooperative Observer Program Network²². Sites are chosen for inclusion in the USHCN according to their spatial coverage, record length, data completeness, and historical stability. USHCN records are subject to rigorous quality control checks and have been demonstrated to be less error-prone than the GSOD³³. We included in the study only those USHCN stations for which at least 90% of daily precipitation data were available in no fewer than 95% of the 480 months of our study period, reducing the total number of stations in the analysis from 1,218 to $n = 601$. This reduced the share of days in the USHCN data set with missing precipitation data to 1.2%. Because any time-dependent missing daily precipitation data could potentially affect our measurements of annual total precipitation and precipitation days, we used a procedure for simulating the occurrence of precipitation on missing data days that has been employed in leading research on over-time precipitation trends^{34,35}. All simulations were conducted at the station-by-month level. For any day with missing precipitation data, we first employed a random-number generator to simulate whether precipitation occurred by using the observed frequency of precipitation within the station-month over the 40-year period of our analysis. We fitted a separate gamma distribution—which has been shown to realistically represent precipitation processes—to each

station's daily precipitation by month (for a total of $601 \times 12 = 7,212$ distributions), using only months for which the station had complete data. A random draw from the fitted distribution was then used to simulate missing daily precipitation for any day in the station-month on which precipitation was simulated to occur. As a robustness check, we carried out the same analysis using GSOD precipitation data; results were similar (Extended Data Table 4c).

Linking weather to population. To estimate population exposure to weather conditions, we used a method employed by health geographers that weights weather station observations based on their distance from population centroids of US counties^{36,37}. Unlike other geographic units we might use to measure population exposure, counties are the smallest unit of geography for which boundaries remained almost entirely unchanged during the 40-year period. We located the population-weighted centroid for each county using block group population and boundary data from the 1990 Census, which was conducted approximately at the midpoint of our study period. We then assigned weights to GSOD and USHCN weather stations located within 160 km of a county's population centroid based on the inverse of the station's squared distance from the centroid. Counties with no weather station within 160 km from their centroids ($n = 66$ of the 3,103 counties in the contiguous United States) were dropped from the analysis; the counties remaining accounted for 98% of the 2010 contiguous US population. The median number of GSOD weather stations assigned to counties was 7; the median number of USHCN stations was 13 (Supplementary Table 1). Extended Data Figure 1 shows a map of the weather stations and counties in our data set.

Our findings are robust to other methods of matching weather conditions to the population. The results were similar when including only stations located within 80 km of population centroids (Extended Data Table 4d). In a separate analysis, we created a Voronoi polygon around each GSOD weather station with valid temperature, humidity and precipitation data (Extended Data Fig. 3) and then assigned population to the polygons by 1990 Census block groups. Using this method—which relies only on a single weather station's data for each block group and does not include USHCN data—we find results similar to those in our main analyses (Extended Data Table 5).

We used daily data to calculate monthly averages by weather station for each of the weather indicators in our data set, yielding data at the station \times year \times month level, and then calculated annual values of January average daily maximum temperature, amount of precipitation, and number of days on which precipitation occurred. We used standard formulas to calculate July average daily mean relative humidity³⁸ and July daily heat index³⁹.

Because we were interested in Americans' experience with the weather rather than distinguishing between short-term natural variability and long-term climate trends, we did not adjust the data to remove urban heat island effects. We also did not adjust for changes in instrumentation or observation routine. Research on the effects of these changes on temperature measurements suggests that the effects are modest and should bias results against our findings. The transition from afternoon to morning temperature observations and the adoption of electronic instruments both had the effect of recording lower maximum temperatures^{4,40,41}. These effects do not seem to vary between winter and summer⁴¹. Because warming that has occurred over the last 40 years has been more pronounced and widespread in winter than in summer, instrument changes would result in understating the amount of January warming that has occurred and the corresponding increase in WPI. The effects of instrument changes on dewpoint temperature, and thus relative humidity measurements, are less systematic over time, but they have been detected at only a small proportion of stations²⁴. Considering the modest role that relative humidity plays in our preference model and the limited evidence that instrumentation substantially affects measurements, we determined that data adjustments were not necessary.

We assigned weather station data to counties using the inverse-distance weights described earlier. Because our interest was in population exposure, we weighted our annual indicators by 2010 county population⁴². We used constant population weights—rather than adjusting them for shifts over time in county population—to isolate the impact of weather trends from any changes in aggregate exposure that are attributable to population migration or growth. As shown in Table 1, using population weights from 1970 rather than 2010 or unweighted values produced similar results. Summary statistics for the county-level weather indicators over our 40-year study period are reported in Extended Data Table 1; mean values of WPI by county are shown in Fig. 2a.

WPI measures. To produce the results reported in the paper, we used a WPI derived from a population growth model reported in a widely cited study¹⁴ (reported in ref. 14 table 3, model 6). The model includes both linear and quadratic weather terms to flexibly assess preferences about weather. In the model used here, county population growth from 1970 to 2000 is a function of five long-term normal weather indicators: January average daily maximum temperature

(JAN_MAX); July daily heat index (JULY_HI); July average daily mean relative humidity (JULY_RH); annual precipitation (PRECIP_IN); and the number of days on which precipitation occurs annually (PRECIP_DAYS). Control variables include county geographical, coastline and topological features, baseline population density and total population, and baseline shares of county population employed in different sectors, including those tied closely to weather such as agriculture and transportation. Taking the reported coefficients estimating the partial relationships between the weather indicators included in the model and population growth, we calculated a WPI score for each county j in each year t using

$$\begin{aligned} \text{WPI}_{jt} = & 0.0488 \times \text{JAN_MAX}_{jt} + 0.0013 \times \text{JAN_MAX}_{jt}^2 - 0.0170 \times \text{JULY_HI}_{jt} \\ & - 0.0008 \times \text{JULY_HI}_{jt}^2 - 0.0385 \times \text{JULY_RH}_{jt} - 0.0003 \times \text{JULY_RH}_{jt}^2 \\ & - 0.0048 \times \text{PRECIP_IN}_{jt} + 0.0002 \times \text{PRECIP_IN}_{jt}^2 \\ & + 0.0065 \times \text{PRECIP_DAYS}_{jt} - 0.0002 \times \text{PRECIP_DAYS}_{jt}^2 \end{aligned} \quad (1)$$

All weather indicators are centred at their means, and thus the linear term coefficients can be interpreted as the effect of a one-unit shift in the indicator on WPI at the indicator's mean value. As shown in equation (1), the analyses in this study (and all published studies from which we derive WPIs) were conducted using US conventional (imperial) units of measure. To comport with these studies, we employed imperial units in our calculations of all WPIs and then transformed results into SI units to report temperature and precipitation trends.

We checked the robustness of our finding by calculating alternative WPIs based on five other published analyses^{9,11–13} estimating the effect of climate amenities on local population growth. These studies employ simpler treatments of climate amenities, in some cases including only two or three indicators related to temperature, precipitation or humidity. All treat summer and winter temperatures separately. For each study, we developed a WPI based on reported coefficients on the study's weather-related variables (Extended Data Tables 2 and 3). We then used our county-level weather data to calculate annual WPI scores for all US counties from each of these WPI formulas. We were able to measure all weather-related variables at the county level over the entire 40 years except for sunshine hours, a variable that appears in two of the models^{12,13}. Our estimates therefore assume no long-term change in the amount of sunshine experienced by individual counties.

Future warming projections. County-level temperature estimates for the RCP4.5 and RCP8.5 emissions scenarios came from the NCCV, which uses NASA Earth Exchange Downscaled Climate Projections (NEX-DCP30) data to project future changes in climate and water balance for states, counties and hydrologic units^{26,27}. The NEX-DCP30 data set statistically downscales projections from 33 models included in the 5th Climate Model Intercomparison Program (CMIP5) to an 800-m grid. The NCCV includes 30 of the 33 models that cover both emissions scenarios and creates area-weighted averages at the county level. Consistent with the NCCV's presentation of these data, we have examined projections over three time periods—2025–2049, 2050–2074, and 2075–2099—under the two emissions scenarios, comparing mean WPI values within each time period to the observed 1974–2013 means for every county (Extended Data Table 6). Because of data availability, we estimated changes in WPI based only on changes in summer and winter temperatures, weighting counties by their 2010 populations and fixing other weather indicators at their means for the final 10 years of our study

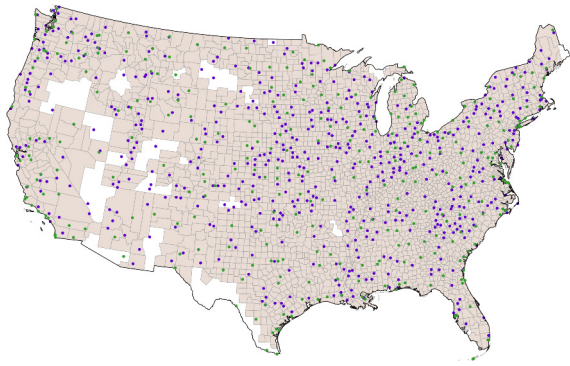
period. Consistent with studies of CMIP5 model performance, we found discrepancies between the observed temperature record and hindcasts yielded by climate models^{7,43}, with the effect that simulated temperature data for the 40-year historical period of our study produce average WPI scores that are lower than those calculated using observed data. Recognizing that discrepancy between modelled and observed temperatures may persist into the future, we performed an additional analysis in which we regression-adjusted projected future WPI scores under all time frames and scenarios to account for the discrepancy. The adjustment was performed by regressing observed annual WPI on simulated WPI derived from the CMIP5 models for the 1974–2013 period, yielding

$$\text{WPI}_{\text{observed}} = 0.691 + 0.463 \times \text{WPI}_{\text{simulated}} \quad (2)$$

We adjusted the projections in Fig. 3a using predictions from this model and display the adjusted projections in Fig. 3b.

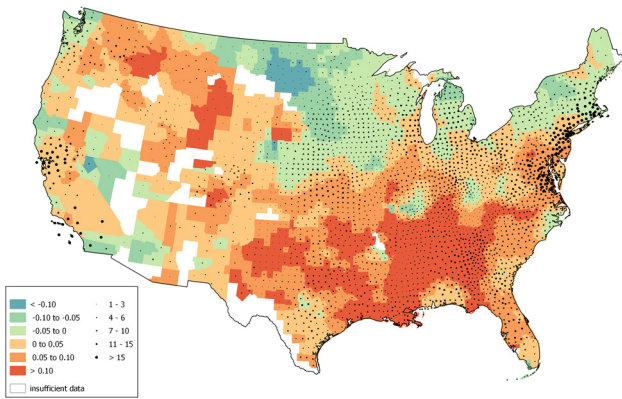
Northern Hemisphere temperate zone projections. We obtained these projections using the Royal Netherlands Meteorological Institute's Climate Change Atlas, which provides CMIP5 climate model output for a variety of countries, seasons, time periods and scenarios through its web-based interface³². For each country, we obtained the mean surface-averaged projections of change in maximum winter and summer near-surface temperatures in the 2075–2099 period under RCP4.5 and RCP8.5 with respect to the reported mean of those observed in the 1974–2013 period (Extended Data Table 7).

31. Minnesota Population Center. *National Historical Geographic Information System: Version 2.0* (Univ. of Minnesota, 2011).
32. Dutch Ministry of Infrastructure and Environment. *KNMI Climate Change Atlas* (Royal Netherlands Meteorological Institute, 2015).
33. New, M., Todd, M., Hulme, M. & Jones, P. Precipitation measurements and trends in the twentieth century. *Int. J. Climatol.* **21**, 1889–1922 (2001).
34. Karl, T. R., Knight, R. W. & Plummer, N. Trends in high-frequency climate variability in the twentieth century. *Nature* **377**, 217–220 (1995).
35. Karl, T. R. & Knight, R. W. Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Am. Meteorol. Soc.* **79**, 231–241 (1998).
36. Hanigan, I., Hall, G. & Dear, K. B. G. A comparison of methods for calculating population exposure estimates of daily weather for health research. *Int. J. Health Geogr.* **5**, 38 (2006).
37. Barreca, A. I. Climate change, humidity, and mortality in the United States. *J. Environ. Econ. Manage.* **63**, 19–34 (2012).
38. Wanielista, M., Kersten, R. & Eaglin, R. *Hydrology: Water Quantity and Quality Control* 2nd edn (Wiley, 1997).
39. Stull, R. B. *Meteorology for Scientists and Engineers* 2nd edn (Brooks/Cole, 2000).
40. Hubbard, K. G. & Lin, X. Reexamination of instrument change effects in the U.S. Historical Climatology Network. *Geophys. Res. Lett.* **33**, L15710 (2006).
41. Quayle, R. G., Easterling, D. R., Karl, T. R. & Hughes, P. Y. Effects of recent thermometer changes in the cooperative station network. *Bull. Am. Meteorol. Soc.* **72**, 1718–1723 (1991).
42. Ruggles, S. J. et al. *Integrated Public Use Microdata Series: Version 5.0* (Univ. of Minnesota, 2014).
43. Sheffield, J. et al. North American climate in CMIP5 experiments. Part I: evaluation of historical simulations of continental and regional climatology. *J. Clim.* **26**, 9209–9245 (2013).

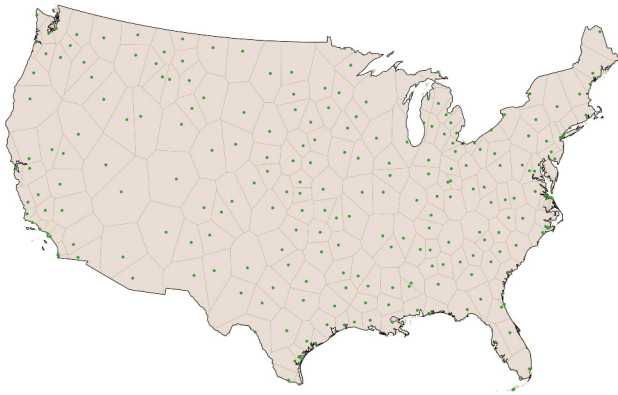


Extended Data Figure 1 | Stations and counties in weather data set.

Green markers indicate GSOD weather stations (temperature and humidity data; $n = 324$). Purple markers indicate USHCN weather stations (precipitation data; $n = 601$). Shaded counties are those with valid weather data ($n = 3,037$). Only counties with population centroids within 160 km of at least one GSOD station and one USHCN station, both reporting valid data, were included in the analysis.



Extended Data Figure 2 | Change in WPI and GSOD weather stations per county. Map shows population growth rate equivalent change in WPI by decade (same as Fig. 2b). Size of dots represents the number of GSOD weather stations assigned to each county.



Extended Data Figure 3 | Voronoi analysis. Map shows locations of GSOD weather stations ($n = 233$) and associated polygons used in Voronoi analysis (Extended Data Table 5).

Extended Data Table 1 | Annual county weather indicators, 1974–2013: summary statistics

Indicator	Mean	Std. dev.	Min.	Max.
Max daily January temp (°C)	7.0	7.4	-17.1	27.9
Max daily July temp (°C)	31.7	3.0	15.7	45.0
July relative humidity (%)	65.4	11.4	15.9	95.4
Max daily July heat index (°C equiv)	39.2	5.9	24.6	57.4
Annual precipitation (mm)	990.4	382.6	36.2	2,786.6
Annual precipitation (days)	104.8	27.1	10.7	210.6
Weather preference index score (annual population growth rate equiv.)	0.9	0.7	-1.2	4.5

n = 3,037 counties × 40 years = 121,480 observations. Std. dev., standard deviation.

Extended Data Table 2 | Replication of main results with alternative WPIs

Change in WPI per decade, 1974-2013 Population growth or migration rate equivalent								
Key to notes in Table 3	Source of estimates	percentiles			un-weighted	means		% experiencing increase in WPI, weighted by 2010 population
		25 th	50 th	75 th		weighted by 1970 population	weighted by 2010 population	
A	(9) (city estimates)	0.03	0.04	0.05	0.04	0.04	0.04	95.6%
B	(9) (MSA estimates)	0.03	0.04	0.05	0.04	0.04	0.04	96.1%
C	(11)	0.07	0.10	0.13	0.10	0.08	0.08	95.8%
D	(12) *	0.09	0.13	0.17	0.13	0.10	0.09	93.3%
E	(13)	0.0003	0.0013	0.0029	0.0016	0.0010	0.0009	72.3%
	(14)	0.0005	0.0482	0.0922	0.0459	0.0401	0.0432	79.6%

Shown are the statistics in the final row of Table 1 as calculated with WPIs derived from alternative population growth studies. Detailed descriptions of the studies appear in Extended Data Table 3. Main results derived from estimates reported previously¹⁴ are repeated here for ease of comparison.

*Expressed in units of the standard deviation of logged population growth. See Extended Data Table 3 for details.

Extended Data Table 3 | Description of estimates of alternate WPIs

Key to results in Table 2	Source study, geographic units, dependent variable (unit of measurement)	Time span of source study	Weather variables used to estimate WPI	Controls	Notes on calculation of WPI
A	(9) U.S. cities, growth (annual %)	1980-2000	(1) (2) (3)	Baseline population, population density, public transit use, age, manufacturing employment, region fixed effects	Calculated using an average of coefficients estimated for 1980-90 and 1990-2000. (Table 6, Models 4 and 8, p.155)
B	(9) U.S. MSAs, growth (annual %)	1980-2000	(1) (2) (3)	Baseline population, population density, public transit use, age, manufacturing employment, region fixed effects	Calculated using an average of coefficients estimated for 1980-90 and 1990-2000. (Table 7, Models 4 and 8, p.156)
C	(11) U.S. county aggregates, net migration (annual %)	1950-1980	(1) (2)	Baseline population density, % black, education, manufacturing and agriculture employment, census division fixed effects	Calculated using an average of coefficients estimated for 1950-60, 1960-70, and 1970-1980. (Table 1, Models 1, 2 and 3, p.190)
D	(12) U.S. non-metro counties, growth (std. dev. of logged annual %)	1970-1996	(1) (4) (5)	Water area, topography, county economic type, poverty, urbanicity, days of sunshine in January	Calculated using standardized regression coefficients, the only estimates reported by the study. (Table 4, Model 4, p. 8)
E	(13) U.S. counties, growth (annual %)	1950-2000	(1) (2) (4)	Water area, distance to coastline, topography, distance from metropolitan center, state fixed effects	Calculated using a weighted average of coefficient estimates for non-metro, small metro, and large metro areas. Weights correspond to % of U.S. population in these three types of areas in 1990 (.16, .11, and .71 respectively). (Table 1, Models 1, 2 and 3, p. 531)

Weather variables: (1) average daily January temperature; (2) average daily July temperature; (3) average annual precipitation (inches); (4) July relative humidity; (5) summer temperateness, derived from (negative) residuals of regression of average daily July temperature on average January temperature.

Extended Data Table 4 | Robustness checks

Indicator	percentiles			un-weighted	means		% experiencing increase, weighted by 2010 population
	25 th	50 th	75 th		weighted by 1970 population	weighted by 2010 population	
a. Raise GSOD weather-station missing-data threshold from 50% to 75% (<i>n</i> = 3,008 counties; 290 GSOD stations)							
Max daily January temp (°C)	0.55	0.70	0.85	0.69	0.62	0.58	99.9%
Max daily July temp (°C)	-0.12	0.02	0.16	0.03	0.06	0.07	59.4%
July relative humidity (%)	-0.65	-0.14	0.41	-0.06	-0.18	-0.20	36.7%
Max daily July heat index (°C equiv)	-0.25	0.03	0.29	0.02	0.06	0.07	55.4%
Annual precipitation (mm)	-12.27	8.01	22.59	5.37	10.64	5.64	60.2%
Annual precipitation (days)	-0.99	0.38	1.70	0.31	0.73	0.31	63.9%
Weather preference index (WPI) score	0.00	0.05	0.10	0.05	0.04	0.05	80.5%
b. Limit analysis to GSOD stations that did not change location from 1974 to 2013 (<i>n</i> = 2,998 counties; 290 GSOD stations)							
Max daily January temp (°C)	0.54	0.70	0.84	0.69	0.61	0.58	99.9%
Max daily July temp (°C)	-0.13	0.02	0.15	0.03	0.06	0.07	59.3%
July relative humidity (%)	-0.65	-0.19	0.33	-0.06	-0.17	-0.20	32.9%
Max daily July heat index (°C equiv)	-0.24	0.02	0.27	0.01	0.06	0.06	56.2%
Annual precipitation (mm)	-12.50	7.74	22.59	5.28	10.63	5.62	60.1%
Annual precipitation (days)	-0.98	0.38	1.70	0.31	0.73	0.31	63.8%
Weather preference index (WPI) score	0.01	0.05	0.10	0.05	0.04	0.05	81.3%
c. Use GSOD precipitation data (<i>n</i> = 2,933 counties; 233 GSOD stations)							
Max daily January temp (°C)	0.54	0.69	0.83	0.68	0.61	0.58	99.9%
Max daily July temp (°C)	-0.13	0.02	0.16	0.03	0.05	0.06	59.4%
July relative humidity (%)	-0.69	-0.24	0.40	-0.09	-0.23	-0.24	31.0%
Max daily July heat index (°C equiv)	-0.27	0.01	0.28	0.00	0.04	0.05	50.5%
Annual precipitation (mm)	-30.74	-12.66	4.17	-16.36	-10.97	-13.67	37.0%
Annual precipitation (days)	-2.19	0.18	2.83	0.41	-0.63	-0.80	45.8%
Weather preference index (WPI) score	0.03	0.07	0.12	0.06	0.05	0.05	81.7%
d. Reduce maximum distance between county centroids and weather stations from 160 to 80 km (<i>n</i> = 1,828 counties; 322 GSOD stations; 596 USHCN stations)							
Max daily January temp (°C)	0.52	0.70	0.84	0.68	0.61	0.59	97.7%
Max daily July temp (°C)	-0.15	0.01	0.18	0.02	0.07	0.07	61.8%
July relative humidity (%)	-0.78	-0.08	0.68	0.00	-0.16	-0.18	40.2%
Max daily July heat index (°C equiv)	-0.25	0.03	0.36	0.02	0.08	0.08	56.0%
Annual precipitation (mm)	-9.47	10.56	28.36	8.57	12.25	7.51	63.0%
Annual precipitation (days)	-1.26	0.64	2.43	0.58	0.72	0.39	57.0%
Weather preference index (WPI) score	-0.01	0.04	0.10	0.04	0.04	0.04	75.3%

Shown are the statistics in Table 1 as calculated with various weather data measurement approaches.

Extended Data Table 5 | Robustness check: assign population to weather stations using Voronoi polygons

Indicator	<u>percentiles</u>			<u>means</u>		% experiencing increase, weighted by 1990 population
	25 th	50 th	75 th	un-weighted	weighted by 1990 population	
Max daily January temp (°C)	0.39	0.62	0.82	0.59	0.58	97.7%
Max daily July temp (°C)	-0.15	0.05	0.26	0.08	0.06	59.6%
July relative humidity (%)	-0.91	-0.24	0.59	-0.12	-0.26	40.2%
Max daily July heat index (°C equiv)	-0.23	0.06	0.46	0.08	0.04	48.5%
Annual precipitation (mm)	-34.80	-16.38	5.23	-18.96	-13.46	31.3%
Annual precipitation (days)	-3.70	-0.36	4.71	0.44	-0.81	32.4%
Weather preference index (WPI) score	-0.01	0.07	0.13	0.06	0.08	76.9%

Shown are the statistics in Table 1 calculated with temperature, humidity and precipitation data from GSOD weather stations and population assigned to stations using 233 Voronoi polygons.

Extended Data Table 6 | Projections of change in WPI scores under two RCP scenarios, 2025–2099

Change in county WPI score from 1974-2013 means						
Scenario and time period	percentiles			means		% experiencing decrease in WPI score compared to 1974-2013, weighted by 2010 population
	25 th	50 th	75 th	un-weighted	weighted by 2010 population	
RCP4.5:						
2025-2049	-0.17	-0.05	0.03	-0.10	-0.15	59.7%
2050-2074	-0.28	-0.10	-0.01	-0.17	-0.20	68.3%
2075-2099	-0.37	-0.14	-0.03	-0.22	-0.23	71.1%
RCP8.5:						
2025-2049	-0.21	-0.07	0.01	-0.13	-0.17	66.1%
2050-2074	-0.61	-0.35	-0.12	-0.39	-0.35	82.0%
2075-2099	-1.20	-0.77	-0.34	-0.79	-0.65	87.8%

Shown are summary statistics of the projected within-county change in WPI scores relative to their 1974–2013 means across 3,037 counties under two scenarios (RCP4.5 and RCP8.5).

Extended Data Table 7 | Projected change in maximum winter and summer temperatures in major regions and countries of the temperate zone of the Northern Hemisphere under two RCP scenarios

Projected change in maximum near-surface seasonal temperatures (°C)						
Country or region	RCP 4.5			RCP 8.5		
	Summer	Winter	Difference	Summer	Winter	Difference
United States	2.74	2.33	0.41	4.93	4.02	0.91
Canada	2.45	3.99	-1.54	4.60	7.06	-2.46
Europe	2.79	2.47	0.32	5.15	4.48	0.67
Russia	2.56	3.55	-0.99	4.70	6.59	-1.89
China	2.38	2.55	-0.17	4.36	4.56	-0.20
Japan	2.09	2.20	-0.11	3.76	3.89	-0.13

Shown are means of surface-averaged projections of change in maximum winter and summer temperatures with respect to the means of those observed in the 1974–2013 period. Projections are derived under two scenarios (RCP4.5 and RCP8.5) for the time period 2075–2099. Summer is defined as June, July and August; winter as December, January and February.

The Parkfield tremors reveal slow and fast ruptures on the same asperity

Deepa Mele Veedu¹ & Sylvain Barbot¹

The deep extension of the San Andreas Fault is believed to be creeping, but the recent observations of tectonic tremors from these depths indicate a complex deformation style¹. In particular, an isolated tremor source near Parkfield has been producing a sequence of low-frequency earthquakes² that indicates an uncommon mechanism of stress accumulation and release. The tremor pattern regularly oscillated between three and six days from mid-2003 until it was disrupted by the 2004 magnitude 6.0 Parkfield earthquake. After that event, the tremor source ruptured only about every three days, but over the next two years it gradually returned to its initial alternating recurrence pattern. The mechanism that drives this recurrence pattern is unknown. Here we use physics-based models to show that the same tremor asperity—the region from which the low-frequency earthquakes radiate—can regularly slip in slow and fast ruptures, naturally resulting in recurrence intervals alternating between three and six days. This unusual slip behaviour occurs when the tremor asperity size is close to the critical nucleation size of earthquakes. We also show that changes in pore pressure following the Parkfield earthquake can explain the sudden change and gradual recovery of the recurrence intervals. Our findings suggest a framework for fault deformation in which the same asperity can release tectonic stress through both slow and fast ruptures.

The San Andreas Fault is one of the few strike-slip faults in the world where tectonic tremors are observed, and they emanate from its predominantly creeping depths (Fig. 1). Particularly, near Parkfield an isolated tremor source exhibits a striking pattern of recurrence intervals, radiating bursts of low-frequency earthquakes (LFEs) alternately about every three and six days (Fig. 2a, b). This doubling of recurrence intervals is technically referred to as period doubling, and it is generally a sign of the nonlinear dynamics of the mechanical system. A single tremor asperity on a fault isolated from other tremors^{2,3} producing doubled recurrence intervals under constant loading is a peculiar phenomenon that challenges our basic assumptions about how the deep extension of faults deform to release tectonic stress.

The Parkfield tremors may be the best-observed manifestation of period doubling of seismic events in a tectonic setting (Fig. 2a, b). This pattern began in mid-2003 and was sustained for about 18 months until the nearby 2004 moment magnitude (M_w) 6.0 Parkfield earthquake disrupted the tremor recurrence intervals. The Parkfield earthquake introduced overall shorter recurrence intervals and the initial alternating recurrence pattern was recovered after two years, indicating a great sensitivity and a long-lasting response of the tremor source to nearby stress perturbations. The source characteristics of the tremors (the number of LFEs per burst and their amplitude²) can be separated into two types of events based on their recurrence intervals throughout the entire sequence, despite the perturbation (Fig. 2a, b). Simulating the recurrence times and rupture characteristics of earthquakes is an over-arching goal of earthquake science, and represents a significant challenge. In this study, we focus on explaining the general recurrence patterns of the Parkfield tremors. We then discuss the implications of the possible micro-mechanics of LFE generation.

Tremors are generally understood as the seismic manifestation of underlying slow-slip events promoting the rapid failure of micro-asperities^{4–7}, but the specific features of the Parkfield tremors still elude explanation. Slow slip may be an important component of the underlying mechanism, but invoking events with regular slip is insufficient to explain the period-doubling pattern and the systematic changes in source characteristics of the Parkfield tremors. The period-doubling tremor source may represent the closest observation to the natural response of faults at lower-crustal conditions because of its isolation from other tremors. Hints that period doubling can be the natural response of a single patch are found in analogue experiments, where this phenomenon occurs during short-lived transients⁸. But as of now, no consensus exists regarding the mechanism that can explain the recurrence interval pattern of the Parkfield tremors.

Here we present a self-consistent numerical model with hundreds of slip cycles in a single simulation to explain the period-doubling of the Parkfield tremors and their change in behaviour when perturbed by the 2004 Parkfield earthquake. We use a three-dimensional model based on rate- and state-dependent friction, a constitutive law that is adequate to represent frictional resistance at low speed and short seismic slip distance^{9–11}. The model incorporates all stages of earthquake dynamics, including the nucleation, propagation and termination of slip events, and the radiation of seismic waves in the surrounding rocks^{12,13}. We consider a strike-slip fault segment embedded in an elastic medium, incorporating a homogeneous patch that promotes unstable slip surrounded by a region supporting stable slip, the whole system being driven at a constant plate rate at the boundaries (Extended Data Fig. 1a, Fig. 2c). The recurrence pattern of the Parkfield tremors is assumed to be due to the spontaneous response of a single patch, as opposed to the complex interactions of adjacent patches. We model the dynamics of the system assuming that the LFEs emanate from micro-asperities scattered in the source region that are triggered during slip episodes, even though these micro-asperities are not explicitly included (Fig. 2c, d).

To explain the period-doubling pattern before the 2004 Parkfield earthquake, we use geometrical and frictional parameters that produce slow and fast ruptures on a single asperity. Fast ruptures typically occur along areas of the fault where the frictional resistance decreases with increased slip velocity, so-called velocity-weakening asperities. Slow and fast ruptures with less and more slip can emerge spontaneously (Fig. 2c, Extended Data Fig. 2c) when R , the radius or half length of the asperity, is close to the theoretical nucleation size h^* , a parameter that depends on friction properties and effective confining stress and that describes the minimum patch size needed to produce a seismic rupture¹⁴. Specifically, for $R = 3$ m and $h^* = 1.74$ m, the fault patch fully ruptures in a period-doubling sequence every three and six days, before the perturbation brought by the 2004 Parkfield earthquake (Extended Data Figs 1b, 2c). This is in contrast with previous studies^{13,15–17} where complex interseismic behaviour was only observed at the boundary of velocity-weakening regions, leaving the conditions for rupture of the entire asperity with period doubling unidentified. Our model shows less variability in recurrence intervals than in the observations,

¹Earth Observatory of Singapore, Asian School of the Environment, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore.

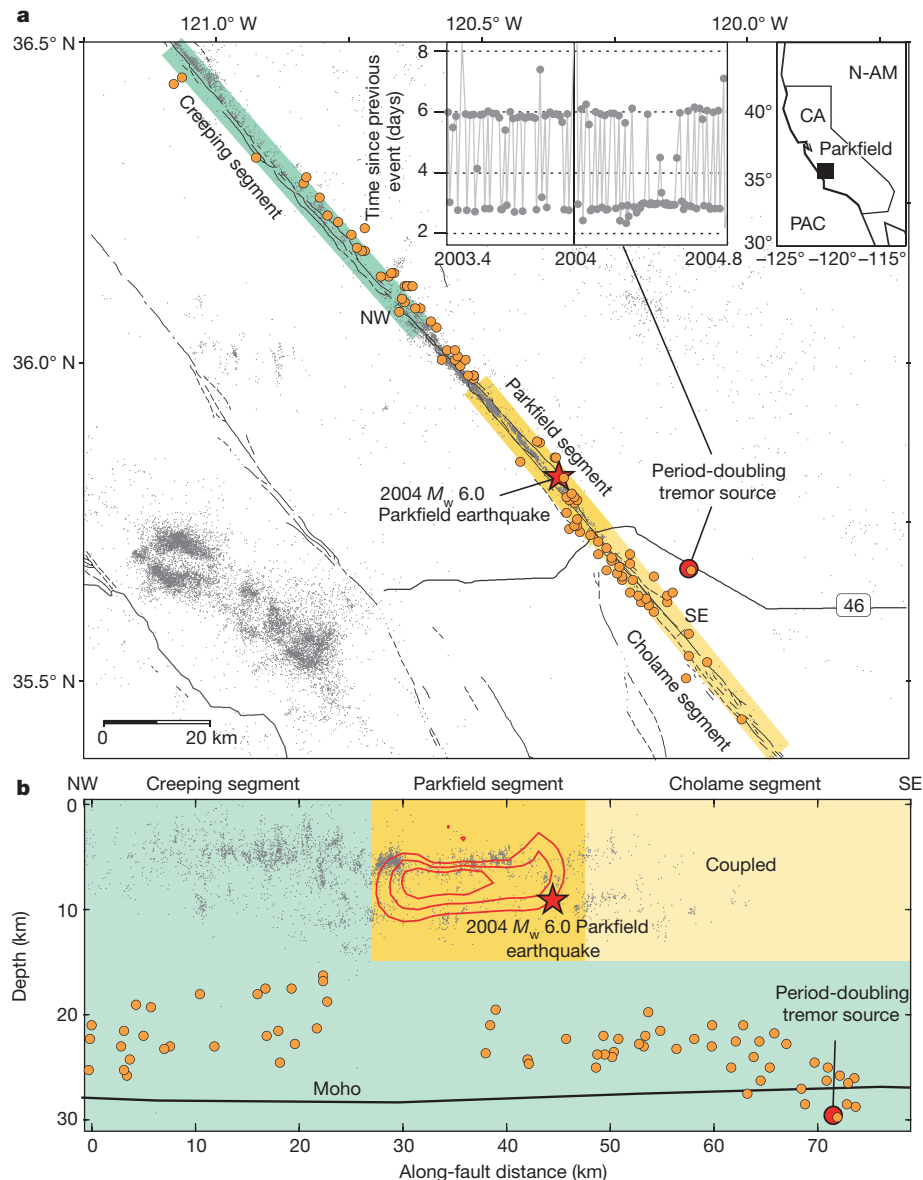


Figure 1 | The Parkfield tremors. **a**, Seismicity map near the Parkfield segment of the San Andreas Fault with tremors from ref. 27 (orange circles) and seismicity from the Northern California Earthquake Data Centre (NCEDEC) catalogue from 2000 to 2004 (grey dots). The red star shows the hypocentre of the 2004 M_w 6.0 Parkfield earthquake, close to Highway 46 (solid line). The period-doubling tremor source (red circle) emits bursts of LFEs alternately about every three and six days (left inset).

Right inset, location of **a** in California (CA), at the boundary between the North American plate (N-AM) and the Pacific plate (PAC). **b**, Fault-parallel cross-section of seismicity; positions of NW and SE are given in **a**. The period-doubling tremor source is around 30 km depth, in the creeping zone, below the locked Cholame segment. The slip distribution of the Parkfield earthquake from ref. 13 is shown with the 20-, 40- and 60-cm contours. The Moho depth is from ref. 28.

presumably due to its overly simple design. Among many unknowns, the geometry of the tremor source and the distribution of frictional parameters may be more complex than we assume. Yet, data and model are sufficiently close to demonstrate the likelihood of a fault deformation where the tectonic stress is released in alternating events with less and more slip on the same patch to result in doubling recurrence intervals. If such behaviour can be reproduced with a relatively simple setup, it is likely to occur spontaneously in nature.

Our numerical investigations indicate that period doubling can occur on faults of almost any size (Extended Data Fig. 3). The period doubling occurs over a wider range of parameters on elongated asperities (Extended Data Fig. 4), which may occur more often in nature than circular or square ones. However, reproducing the three and six days recurrence intervals of the Parkfield tremors requires near-lithostatic pore pressure, in agreement with observations suggesting high sensitivity of tremors to tidal stress and remote triggering^{18–21}. We can

reproduce the three and six day recurrence pattern only if the effective confining pressure, $\bar{\sigma} = \sigma - \alpha p$ (where σ is the lithostatic stress, α is the coefficient of effective stress of order one, and p is the pore pressure), is lower than 50 MPa. For example, with a confining pressure of 5 MPa the asperity size is $R = 33$ m and the slip events have a moment magnitude close to 1 (Extended Data Fig. 5); for a confining pressure of 50 MPa the asperity size is $R = 3$ m and it produces slip events of moment magnitude below 0 (Fig. 2c, d). Currently, there is no direct estimate of the moment magnitude of the Parkfield tremors. Given the advanced and highly tuned techniques used to separate them from seismic noise^{2,3}, it is possible that their magnitude is in the range of our models.

The tremor sequence contains a two-year transient following the 2004 M_w 6.0 Parkfield earthquake that provides an opportunity to test the model with perturbations. The sudden disappearance of the six-day recurrence intervals and period multiplying after the Parkfield

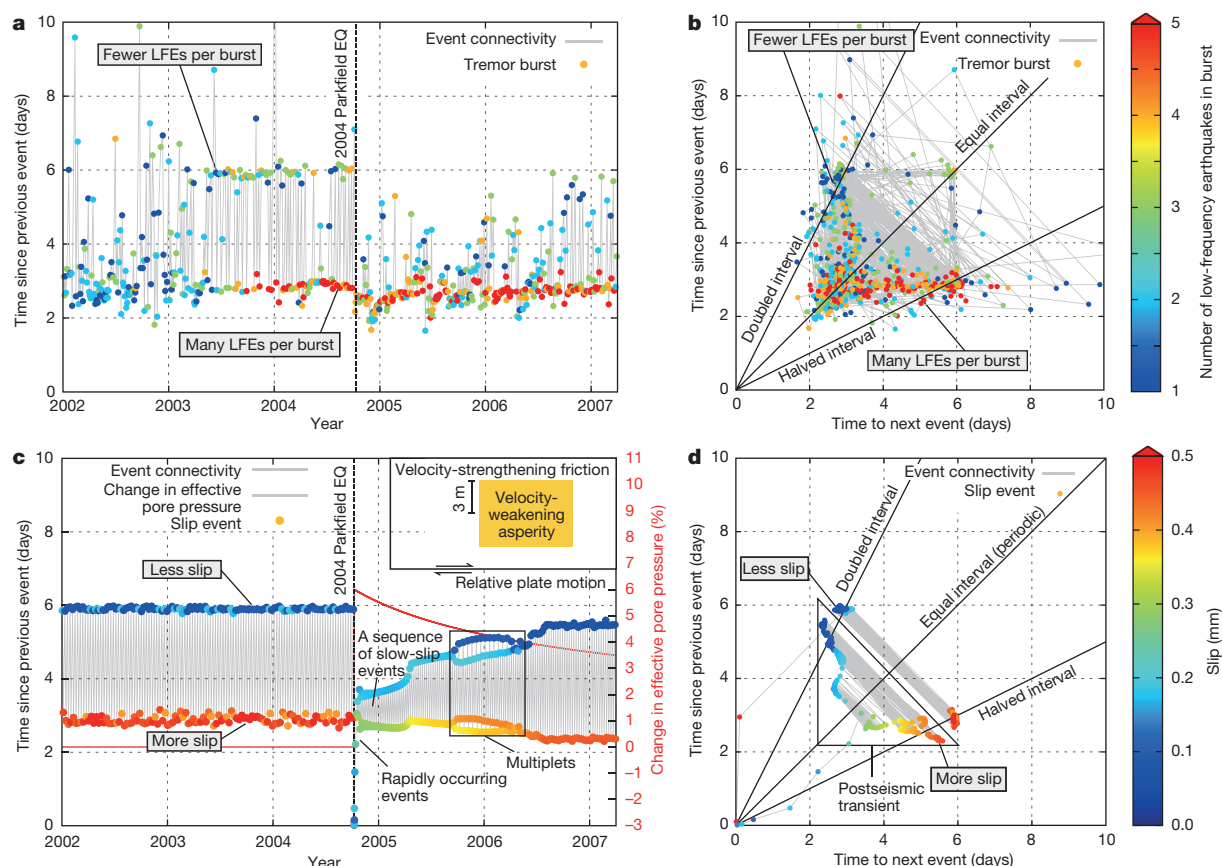


Figure 2 | Recurrence pattern and source characteristics of the period-doubling Parkfield tremors and numerical simulations. **a**, Observed tremor recurrence pattern and number of LFEs per burst, modified from ref. 2. Each tremor burst consists of 1 to 5 LFEs in most cases. Bursts preceded by a shorter recurrence interval are associated with more LFEs throughout the sequence (consecutive bursts are connected with a grey line). The regular oscillation of recurrence intervals between about three and six days is from mid-2003 to the onset of the 2004 Parkfield earthquake. The sudden disappearance of the 6-day recurrence intervals coincides with the 2004 Parkfield earthquake. The gradual recovery of recurrence intervals is from the onset of the Parkfield earthquake to the end of 2007, around two years. **b**, Characterization of the dynamics of the tremors. Most events falling on the equal pre- and post-intervals (diagonal line) occur at the onset of the 2004 Parkfield earthquake, when period

doubling is momentarily interrupted. The pairs of period-doubling events fall on the doubled-interval and halved-interval lines. The colour denotes the number of LFEs in the burst. **c**, Numerical simulation of the Parkfield tremor activity incorporating a change in effective pore pressure (red profile) after the Parkfield earthquake. The top right inset illustrates the model geometry with $R = 3$ m. The coloured circles denote slip of events. Fast ruptures have more slip and are preceded by shorter recurrence intervals. The 'multiplets' box highlights a period of the simulation with a long sequence of interconnected events. **d**, Characterization of the dynamics of the slip events from numerical simulations. More (less) slip of the fast (slow) ruptures is preserved after the perturbation of the Parkfield earthquake. The colour bar represents slip magnitude of the modelled events in **c** and **d**.

earthquake are probably due to changes in normal stress. The effect is two fold: a sudden pore-pressure increase causes a rapid fire of slip events with shorter recurrence intervals (this is the direct, triggering effect); but this change in normal stress places the patch firmly in the sub-critical regime, promoting a sequence of slow slip events (Fig. 2c, d). The two-year transient can be explained by a decay of the pore-pressure anomaly, during which the fault patch slips in more complex sequences, before returning to a pattern close to the pre-earthquake pattern after two years (Fig. 2c). The amplitude and the timescale of the pore-pressure evolution control the degree of perturbation of the recurrence intervals during the transient, and we choose these values by trial and error to best reproduce the observations. The differences between our models and the original observations are probably due to our simplifying assumptions or missing events due to the difficulty of detection, but it is striking that the number of LFEs and modelled slip events are persistent and follow the same pattern before and after the Parkfield perturbation (Fig. 2, Extended Data Fig. 2).

The postseismic transient of the Parkfield tremors, with the temporary suppression of period doubling, is qualitatively well reproduced assuming the temporal decay of a positive pore-pressure perturbation representing about 6% of the background confining pressure (Fig. 2c).

This denotes either pore fluid diffusion or the damage and recovery of the solid matrix (the pore pressure is related to the fault normal stress through the effective stress coefficient α of order 1). If the background pore pressure is close to the lithostatic stress (about 1 GPa at 30 km depth), only a few tenths of a per cent change in the coefficient of effective stress is sufficient to explain the transient. This is a plausible scenario, as the 2004 Parkfield earthquake did induce a change of shear wave velocity of this order of magnitude in the ambient rocks²².

Equipped with a model that can explain the main features of the Parkfield tremor recurrence patterns, we discuss implications for the mechanics of LFEs at smaller spatial scales than our models can resolve numerically (Fig. 3). As the magnitude of slip of fast ruptures is about twice that of slow ruptures (Fig. 2, Extended Data Fig. 2), the micro-asperities may break more often in this case. But it is unclear whether the observed larger number of LFEs in fast ruptures results from a single micro-asperity rupturing multiple times or from several micro-asperities rupturing sequentially. Every modelled event lasts for a much shorter time than the LFE bursts. And every fast rupture is elasto-dynamic in our simulations, and therefore may be detected seismically. It is possible that LFE generation continues after the underlying slip event and that the fast ruptures represent one of the detected

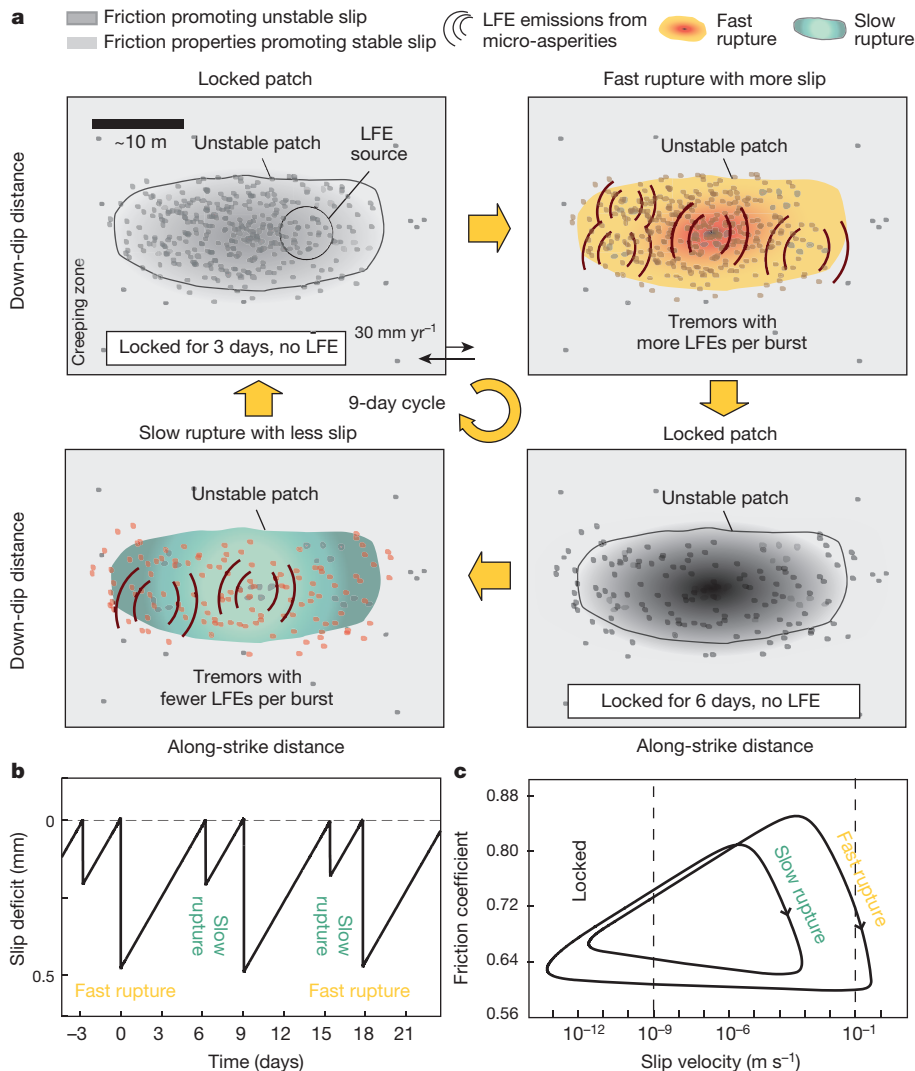


Figure 3 | The modelled mechanics behind the period-doubling tremors. **a**, The tremor source slips in successions of slow and fast ruptures, resulting in two intrinsic recurrence intervals. Fast ruptures (top right panel) produce more slip, resulting in many LFEs, and are followed by a six-day locked period (bottom right panel). Slow ruptures (bottom left panel) produce less slip and result in fewer LFEs, and are followed by a

three-day locked period (top left panel). The full cycle repeats every about nine days. **b**, Slip-deficit evolution with a time-predictable sequence of slip events. **c**, Phase diagrams of period-doubling cycles with slow and fast ruptures. The slip, frictional resistance, and slip velocity are for the centre of the patch. The vertical dashed lines indicate the velocity bounds for a locked asperity and a fast rupture.

LFEs. Alternatively, dilatancy²³, or other additions to the rate-and-state friction framework²⁴ not accounted for in our simulations, would prevent the underlying slip event reaching seismic speed. A remaining puzzling observation is the lower amplitude of those tremors composed of many LFEs². If the area of rupture of micro-asperities varies between slow and fast events due to the unusual loading conditions, it may lead to the observed variability in amplitude.

Our interpretation of the underlying mechanism behind the Parkfield tremors is supported by observations that span hundreds of tremor cycles, an observation window that is not available for great and giant earthquakes. If the mechanism driving sequences of slow and fast ruptures of the same overall patch were behind tremors and earthquakes alike, our results would imply that the occurrence of isolated slow-slip episodes might not necessarily indicate a reduced seismic hazard, as an earthquake could rupture subsequently at the same location.

The 2011 M_w 9.1 Tohoku-Oki and the 2014 M_w 8.1 Iquique earthquakes may have been preceded by slow-slip events with significant overlap with the main rupture^{25,26}, contradicting our general expectations about fault behaviour. It is possible that, as for the Parkfield tremors, this succession of events is promoted by a particular combination of geometric and frictional properties of the seismogenic zone and its surroundings.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 May 2015; accepted 26 January 2016.

Published online 4 April 2016.

- Shelly, D. R. Migrating tremors illuminate complex deformation beneath the seismogenic San Andreas Fault. *Nature* **463**, 648–652 (2010).
- Shelly, D. R. Periodic, chaotic, and doubled earthquake recurrence intervals on the deep San Andreas Fault. *Science* **328**, 1385–1388 (2010).
- Shelly, D. R. Complexity of the deep San Andreas Fault zone defined by cascading tremor. *Nature Geosci.* **8**, 145–151 (2015).
- Ito, Y., Obara, K., Shiomi, K., Sekine, S. & Hirose, H. Slow earthquakes coincident with episodic tremors and slow slip events. *Science* **315**, 503–506 (2007).
- Schwartz, S. Y. & Rokosky, J. M. Slow slip events and seismic tremor at circum-Pacific subduction zones. *Rev. Geophys.* **45**, RG3004 (2007).
- Daub, E. G., Shelly, D. R., Guyer, R. A. & Johnson, P. A. Brittle and ductile friction and the physics of tectonic tremor. *Geophys. Res. Lett.* **38**, L10301 (2011).
- Ghosh, A., Vidale, J. E. & Creager, K. C. Tremor asperities in the transition zone control evolution of slow earthquakes. *J. Geophys. Res.* **117**, B10301 (2012).
- Gu, Y. & Wong, T.-F. in *Nonlinear Dynamics and Predictability of Geophysical Phenomena* (eds Newman, W. I., Gabrieli, A. & Turcotte, D. L.) 15–35 (Geophys. Monogr. 83, AGU, 1994).
- Dieterich, J. H. Modeling of rock friction 1. Experimental results and constitutive equations. *J. Geophys. Res.* **84**, 2161–2168 (1979).
- Scholz, C. H. Earthquakes and friction laws. *Nature* **391**, 37–42 (1998).

11. Lapusta, N. & Barbot, S. in *The Mechanics of Faulting: From Laboratory to Real Earthquakes* (eds Bizzarri, A. & Bhat, H. S.) 153–207 (Research Signpost, 2012).
12. Noda, H. & Lapusta, N. Three-dimensional earthquake sequence simulations with evolving temperature and pore pressure due to shear heating: effect of heterogeneous hydraulic diffusivity. *J. Geophys. Res.* **115**, B12314 (2010).
13. Barbot, S., Lapusta, N. & Avouac, J.-P. Under the hood of the earthquake machine: towards predictive modeling of the seismic cycle. *Science* **336**, 707–710 (2012).
14. Rubin, A. M. & Ampuero, J.-P. Earthquake nucleation on (aging) rate and state faults. *J. Geophys. Res.* **110**, B11312 (2005).
15. Liu, Y. & Rice, J. R. Spontaneous and triggered aseismic deformation transients in a subduction fault model. *J. Geophys. Res.* **112**, B09404 (2007).
16. Matsuzawa, T., Hirose, H., Shibasaki, B. & Obara, K. Modeling short- and long-term slow slip events in the seismic cycles of large subduction earthquakes. *J. Geophys. Res.* **115**, B12301 (2010).
17. Noda, H. & Hori, T. Under what circumstances does a seismogenic patch produce aseismic transients in the later interseismic period? *Geophys. Res. Lett.* **41**, 7477–7484 (2014).
18. Rubinstein, J. L. *et al.* Non-volcanic tremor driven by large transient shear stresses. *Nature* **448**, 579–582 (2007).
19. Peng, Z. *et al.* Strong tremor near Parkfield, CA, excited by the 2002 Denali Fault earthquake. *Geophys. Res. Lett.* **35**, L23305 (2008).
20. Thomas, A. M., Nadeau, R. M. & Bürgmann, R. Tremor-tide correlations and near-lithostatic pore pressure on the deep San Andreas Fault. *Nature* **462**, 1048–1051 (2009).
21. Ide, S. Striations, duration, migration and tidal response in deep tremor. *Nature* **466**, 356–359 (2010).
22. Brenguier, F. *et al.* Postseismic relaxation along the San Andreas Fault at Parkfield from continuous seismological observations. *Science* **321**, 1478–1481 (2008).
23. Segall, P., Rubin, A. M., Bradley, A. M. & Rice, J. R. Dilatant strengthening as a mechanism for slow slip events. *J. Geophys. Res.* **115**, B12305 (2010).
24. Rubin, A. M. Designer friction laws for bimodal slow slip propagation speeds. *Geochem. Geophys. Geosyst.* **12**, Q04007 (2011).
25. Ito, Y. *et al.* Episodic slow slip events in the Japan subduction zone before the 2011 Tohoku-Oki earthquake. *Tectonophysics* **600**, 14–26 (2013).
26. Meng, L., Huang, H., Bürgmann, R., Ampuero, J.-P. & Strader, A. Dual megathrust slip behaviors of the 2014 Iquique earthquake sequence. *Earth Planet. Sci. Lett.* **411**, 177–187 (2015).
27. Shelly, D. R. & Hardebeck, J. L. Precise tremor source locations and amplitude variations along the lower-crustal central San Andreas Fault. *Geophys. Res. Lett.* **37**, L14301 (2010).
28. Tape, C., Plesch, A., Shaw, J. H. & Gilbert, H. Estimating a continuous Moho surface from the California Unified Velocity Model. *Seismol. Res. Lett.* **83**, 728–735 (2012).

Acknowledgements We thank D. R. Shelly for sharing his tremor data and J.-P. Avouac for his comments on an earlier version of the manuscript. This work comprises Earth Observatory of Singapore contribution no. 115. S.B. is funded by the National Research Foundation Singapore under its Singapore NRF Fellowship scheme (National Research Fellow Award no. NRF-NRFF2013-04), by the Earth Observatory of Singapore, by the National Research Foundation Singapore and by the Singapore Ministry of Education under the Research Centres of Excellence initiative.

Author Contributions D.M.V. and S.B. designed the experiment, analysed the data, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.B. (sbarbot@ntu.edu.sg) or D.M.V. (deepa004@e.ntu.edu.sg).

METHODS

Simulating the Parkfield tremors. The Parkfield tremors, with their unusual recurrence interval pattern, occurred on the San Andreas Fault at around 30 km depth². Here we describe the methodology we employ to simulate this behaviour using a three-dimensional fully dynamic model based on the rate-and state-friction law^{12,29}.

We use the BICycle (Boundary Integral Cycle of Earthquakes) programme, a boundary-integral method to simulate fault slip evolution during all stages of the earthquake cycle, including the nucleation, propagation and termination of aseismic and seismic events^{12,29}. The method has been validated during the Southern California Earthquake Centre (SCEC)/United States Geological Survey (USGS) rupture code verification exercise³⁰. The model has been used to understand the sequence of the M_w 6.0 earthquakes on the Parkfield segment of the San Andreas Fault¹³, as well as nearby repeating earthquakes³¹, and to study the segmentation of subduction-zone earthquakes³². Our model explains the period doublings and halvings with recurrence intervals of three and six days, and the postseismic transient.

Modelling assumptions. Our model has a strike-slip fault segment surrounded by an elastic medium, driven by long-term slip velocity. The model has a single asperity at the centre with homogenous frictional properties a , b , μ_0 and L , described below. This model is based on the rate- and state-dependent friction law derived from laboratory experiments^{9,33–35}:

$$\tau = \mu \bar{\sigma} = \left[\mu_0 + a \ln \frac{V}{V_0} + b \ln \frac{\theta V_0}{L} \right] \bar{\sigma} \quad (1)$$

where τ is the shear stress, μ is the coefficient of friction, $\bar{\sigma}$ is the effective confining pressure ($\bar{\sigma} = \sigma - \alpha p$, where σ is the lithostatic stress, α is the coefficient of effective stress, p is the pore pressure), μ_0 is the steady-state friction parameter at reference slip velocity V_0 , a and b are frictional parameters, L is the characteristic weakening distance, and θ is the state variable representing the breaking and the healing of asperities. The state evolution can be described using the ageing law⁹:

$$\dot{\theta} = 1 - \frac{V\theta}{L} \quad (2)$$

We may write the fault strength at steady state as

$$\tau_{ss} = \left[\mu_0 + (a - b) \ln \frac{V}{V_0} \right] \bar{\sigma} \quad (3)$$

where $(a - b)$ affects the dynamics and style of faulting^{10,33–36}. We assign $(a - b) < 0$ in the velocity-weakening asperity, and $(a - b) > 0$ for the rest of the fault region (Extended Data Fig. 1a). The slip dynamics is also controlled by the nucleation size^{9,14,31,33–39}:

$$h^* = \frac{\pi G b L}{2(b - a)^2 \bar{\sigma}} \quad (4)$$

where G is the rigidity. Depending on the ratio of the nucleation size to the asperity size, creep, slow-slip events, bilateral ruptures and unilateral ruptures can be simulated.

In our model, we use a square asperity with a homogeneous velocity-weakening property surrounded by a velocity-strengthening region in an elastic space, loaded by a constant plate rate of $V_{pl} = 3 \text{ cm yr}^{-1}$. Other asperity geometries can also reproduce the period-doubling behaviour, so the choice of a square asperity is for convenience only.

We assume that the seismic emissions of the tremors are caused by slip on micro-asperities due to slip events on a single, much larger, asperity with homogeneous frictional properties^{3–7,40} (Fig. 3). We did not explicitly incorporate the micro-asperities in our model. We assume that the period doubling is caused by a succession of slow and fast ruptures with smaller and larger slip on the same asperity. Here, we assume that any rupture that results in a slip rate greater than a threshold value ($V_{th} = 10^{-4} \text{ m s}^{-1}$) produces LFEs. The threshold velocity is chosen to be lower than the maximum velocity of slow and fast ruptures, yet significantly larger than the plate rate.

Model parameters. Here, we present two end-member models with a single square-shaped homogeneous velocity-weakening asperity. With a dimensionless ratio $R/h^* = 0.73$ (for an asperity size $R = 3 \text{ m}$), and $R/h^* = 0.80$ (for an asperity size $R = 33 \text{ m}$), we simulate the sustained cycles of period doublings and halvings of the Parkfield tremors with recurrence intervals of about three and six days (Extended Data Figs 2c, 5a). In the limit of a relatively large confining pressure of $\bar{\sigma} = 50 \text{ MPa}$, the parameters of the first end-member model are as follows: $R = 3 \text{ m}$, $L = 5 \mu\text{m}$, $h^* = 1.74 \text{ m}$, $a = 10^{-2}$, $b_{VS} = 6 \times 10^{-3}$, $b_{VW} = 14 \times 10^{-3}$, $\mu_0 = 0.65$ and $G = 30 \text{ GPa}$. The model parameters of the second end member, for a lower effective

confining pressure, are as follows: $\bar{\sigma} = 5 \text{ MPa}$, $R = 33 \text{ m}$, $L = 5 \mu\text{m}$, $h^* = 17.4 \text{ m}$, $a = 10^{-2}$, $b_{VS} = 6 \times 10^{-3}$, $b_{VW} = 14 \times 10^{-3}$, $\mu_0 = 0.65$ and $G = 30 \text{ GPa}$. The friction parameters b_{VS} and b_{VW} represent the value of b for the velocity-strengthening and velocity-weakening regions, respectively.

In both cases, the effective confining pressure implies pore pressure near lithostatic stress conditions, in agreement with the high sensitivity of tremors to tidal stress and remote triggering^{18,19,21,41–44}. For confining pressures lower than 5 MPa, our models produce period doubling, but not at three and six days. For confining pressures greater than 50 MPa, the size of the velocity-weakening patch must be reduced greatly, leading to smaller moment magnitudes of the slip events.

Given the complexity of rate-and-state friction faults and the large number of physical parameters, it is possible that period doubling can be obtained for other combinations of frictional parameters and fault geometry. So our proposed range of model parameters is simply to illustrate the non-uniqueness of the model.

Our slow ruptures do not exceed 10^{-3} m s^{-1} . The fast ruptures, however, reach up to 0.6 m s^{-1} for a short duration (Extended Data Fig. 1b). Elasto-dynamic events occur when the slip velocities exceed the critical slip rate V_{dyn} and the direct effect becomes negligible in comparison to the inertial effect¹⁷:

$$V_{dyn} = 2V_s \frac{a\sigma}{G} \quad (5)$$

here V_s is the shear-wave speed ($3,000 \text{ m s}^{-1}$). Fast ruptures are elasto-dynamic and may be detected in seismograms as one of the LFEs.

Effect of geometry and friction properties on slip stability. We simulate period-doubling sequences of slow and fast ruptures along at least eight orders of moment magnitude. We increase the size of a square asperity as well as the fault dimension accordingly, and keep the dimensionless ratio $R/h^* = 0.73$ the same in all simulations. We obtain period-doubling sequences in asperities of different sizes, such as 6 m, 60 m, 600 m and 6 km, indicating that the period-doubling phenomenon is self-similar and scale independent (Extended Data Fig. 3). We obtain pairs of slow and fast ruptures with equivalent geodetic moment magnitude of $M_w - 1.9$ and $M_w - 0.5$, respectively, for a square asperity with $R = 3 \text{ m}$ and $L = 5 \mu\text{m}$ (Extended Data Fig. 3a). The magnitudes are $M_w 0.07$ and $M_w 1.4$ for $R = 30 \text{ m}$ and $L = 0.05 \text{ mm}$ (Extended Data Fig. 3b). For $R = 300 \text{ m}$ and $L = 0.5 \text{ mm}$, we obtain magnitudes of $M_w 2.1$ and $M_w 3.5$ (Extended Data Fig. 3c). Finally, we get $M_w 3.9$ and $M_w 5.4$ for $R = 3 \text{ km}$ and $L = 5 \text{ mm}$ (Extended Data Fig. 3d). We assume the same rheological parameters of $\bar{\sigma} = 50 \text{ MPa}$, $a = 10^{-2}$, the velocity-weakening regions with $(a - b) = -4 \times 10^{-3}$ and the velocity-strengthening region with $(a - b) = +4 \times 10^{-3}$, except for the characteristic length L , which is used to adjust the ratio $R/h^* = 0.73$ within the period-doubling range in all simulations. These explorations suggest that velocity-weakening friction assigned in a square asperity with different sizes may exhibit period doubling at almost all magnitudes.

Non-circular asperities have more potential to produce sequences of period-doubling slow and fast ruptures (Extended Data Fig. 4). Velocity-weakening asperities with elliptical shapes promote period-doubling events more than a circular asperity does, indicating that the geometry of the velocity-weakening asperity is important in enabling period-doubling events. In these experiments, we keep the nucleation size h^* constant, and change the shape of the asperity. The parameters considered here are $\bar{\sigma} = 50 \text{ MPa}$, $a = 10^{-2}$, $L = 5 \mu\text{m}$. The velocity-weakening regions and the velocity-strengthening regions are characterized by $(a - b) = -4 \times 10^{-3}$ and $(a - b) = +4 \times 10^{-3}$, respectively. We assume the same frictional parameters and confining pressure in all simulations. With this choice of parameters, the non-circular asperities produce period-doubling events while the circular patch does not.

Dimensional analysis of recurrence times for slow and fast ruptures. The difference of time interval following slow and fast ruptures can be understood from a dimensional analysis. The total slip of slow ruptures is bounded by the point of neutral velocity dependence, that is, a few multiples of the characteristic weakening distance L . The time following slow ruptures scales with

$$T_r^{\text{slow}} = \frac{L}{V_{pl}} \quad (6)$$

The total slip of elasto-dynamic ruptures is not bounded by weakening distance; rather, it is controlled by the asperity size. As a result, the recurrence time of fast ruptures scales as

$$T_r^{\text{fast}} = \frac{(b - a)\bar{\sigma}}{G} \frac{R}{V_{pl}} \quad (7)$$

where $(b - a)\bar{\sigma}$ controls the stress drop, and $R(b - a)\bar{\sigma}/G$ controls the total slip. For period-multiplying events, $T_r^{\text{fast}} \neq T_r^{\text{slow}}$ and period-multiplying factors are affected by the characteristic weakening distance.

Pore-pressure changes following the 2004 M_w 6.0 Parkfield earthquake. We model the effect of pore pressure evolution numerically using a diffusion model. We use the approach of Noda and Lapusta^{12,45} to model thermal pressurization. We ignore the shear heating effects and only allow diffusion of external pore pressure perturbations applied at prescribed times. The pore-pressure perturbation Δp follows the diffusion equation

$$\Delta \dot{p} = \alpha_{hy} \nabla^2 \Delta p, \quad t > t^{eq} \quad (8)$$

where α_{hy} is the hydraulic diffusivity, with the conditions $\Delta p(t < t^{eq}) = 0$ and $\Delta p(t^{eq})$ given. The amplitude of the pore-pressure perturbation and the hydraulic diffusivity affect the dynamics of the transient and they are chosen by trial and error to best reproduce the seismic observations. We use a pore pressure perturbation of $\Delta p(t^{eq}) = 3$ MPa for the model with $R = 3$ m and $\bar{\sigma} = 50$ MPa, and $\Delta p(t^{eq}) = 300$ kPa for the model with $R = 33$ m and $\bar{\sigma} = 5$ MPa. In both cases, the pore-pressure perturbation represents about 6% of the background effective confining pressure. The change in effective confining pressure and its temporal decay can explain the postseismic transient following the 2004 M_w 6.0 Parkfield earthquake (Fig. 2c, d, Extended Data Fig. 2c, d). The effective confining pressure influences the critical nucleation size (equation (4)). Changes in the critical nucleation size can move the asperity to different stability regimes leading to different dynamics: an increase in pore pressure in the period-multiplying asperity fosters conditional stability and the occurrence of slow-slip events (shorter recurrence intervals); and a decrease in pore pressure promotes conditional instability and the occurrence of bilateral ruptures (longer recurrence intervals). The sudden termination of period doubling after the Parkfield earthquake can therefore be understood as a change of dynamics with a succession of slow-slip events. As the pore pressure perturbation decays, the stability regime converges towards the pre-earthquake conditions and period doubling with slow and fast rupture is obtained again after two years.

The origin of the perturbation could be damage caused by the nearby earthquake or a direct pore-pressure change from the same. Our numerical simulations should be taken to describe a combined effect of damage evolution and pore pressure diffusion. The effective background confining pressure $\bar{\sigma} = \sigma - \alpha p$ depends on the pore pressure p through the coefficient of effective stress α (dimensionless parameter of order 1), so changes in both pore pressure and the effective coupling coefficient can affect the effective confining pressure. If a nearby earthquake compresses the solid matrix around the tremor source, it gives rise to a pore-pressure change

$$\Delta p = -\frac{B}{3} \Delta \sigma_{kk} \quad (9)$$

where $\Delta \sigma_{kk}$ is the change in isotropic stress and B is Skempton's coefficient. If damage is introduced and alters the mechanical coupling between the solid matrix and the pore fluids, the effective coupling coefficient is perturbed by $\Delta \alpha$. Then, the new effective confining pressure becomes:

$$\bar{\sigma} = \sigma - (\alpha p + \alpha \Delta p + \Delta \alpha p + \Delta \alpha \Delta p) \quad (10)$$

The higher-order term $\Delta \alpha \Delta p$ can be neglected. The other terms can be evaluated to establish the likely source mechanism of the perturbation.

We evaluate the change in pore pressure induced by the 2004 M_w 6.0 Parkfield earthquake using the coseismic slip distribution of Barbot *et al.*¹³ and setting $B = 0.5$ for Skempton's coefficient. We use the Relax program⁴⁶ to evaluate the coseismic isotropic stress in the bulk. We find that the 2004 M_w 6.0 Parkfield earthquake produced a direct change of pore pressure of the order of $\Delta p = 0.7$ kPa near the tremor source, a value much smaller than inferred from our model.

We evaluate the contribution of damage as follows. At 30 km depth, overpressurized fluids (pressure greater than hydrostatic equilibrium) are required to maintain a low effective confining pressure. Assuming $\bar{\sigma} = 50$ MPa and $\alpha = 1$, we get a pore pressure of the order of 800 MPa. In this context, perturbations of the coefficient of effective stress of a few tenths of a per cent incur a change in effective confining pressure of the order of $\Delta \alpha p = 1$ MPa, a value in the middle range of our estimates, which range from 300 kPa to 3 MPa.

Given that the Parkfield earthquake was found to induce a shear velocity reduction in the ambient rocks of the order of 0.1% (ref. 22), the effect of damage evolution may be more prevalent than the direct pore-pressure change.

Implications for the physics of low-frequency emissions. The seismic observations show that most events preceded by a short recurrence interval contain more

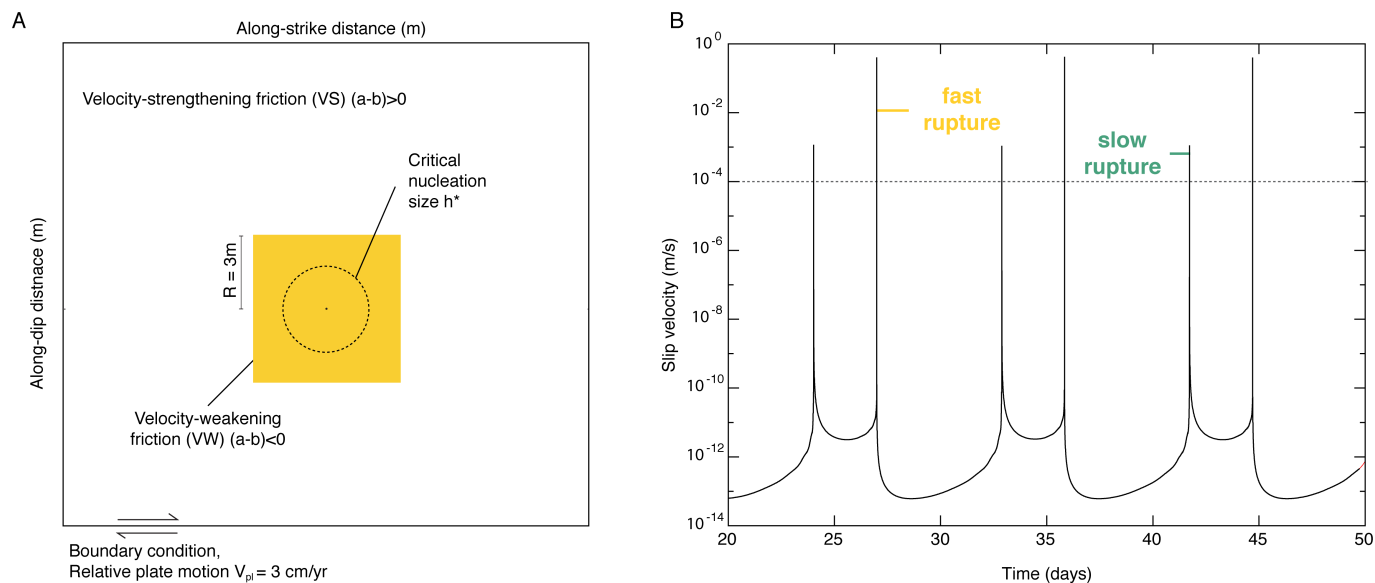
LFEs per burst (Fig. 2a). But the same events have overall less amplitude (ref. 2). The two characteristics are not straightforward to reconcile.

Our physics-based modelling suggests that the unique recurrence pattern of the LFEs at Parkfield may be due to forcing from underlying slip events. In this conception, the LFEs' recurrence pattern mimics the behaviour of the surrounding asperity. But the details of the mechanics of tremor emission are still poorly understood. Because of the difficulty of deciphering tremors from noise^{1-3,27}, some source characteristics such as cumulative moment, moment rate, or stress drop are still unavailable, leaving some aspects of the physical model unconstrained. In particular, it is unknown whether multiple LFEs in a single tremor burst represent multiple emissions from a single micro-asperity or if the source consists of multiple micro-asperities breaking sequentially. The intrinsic response time of the micro-asperity under rapid loading is also poorly characterized, so it is unknown whether slip on micro-asperities is simultaneous with, or comes after, the slip on the underlying patch.

Despite these difficulties, our model makes certain predictions that can be compared to the seismic record. Most importantly, the period-doubling model creates two types of events, that is, the slow and fast ruptures, which are associated with different recurrence intervals. Similarly, the available characteristics of the tremors (number of LFEs per burst, and amplitude) also show two types of events, which can be separated on the basis of the pre or post recurrence intervals (see Fig. 2 and ref. 2). The modelled slip events cannot be the direct representation of the tremor generation process, but it still implies that LFEs are sensitive to the loading rate of the underlying patch.

Sample size. No statistical methods were used to predetermine sample size.

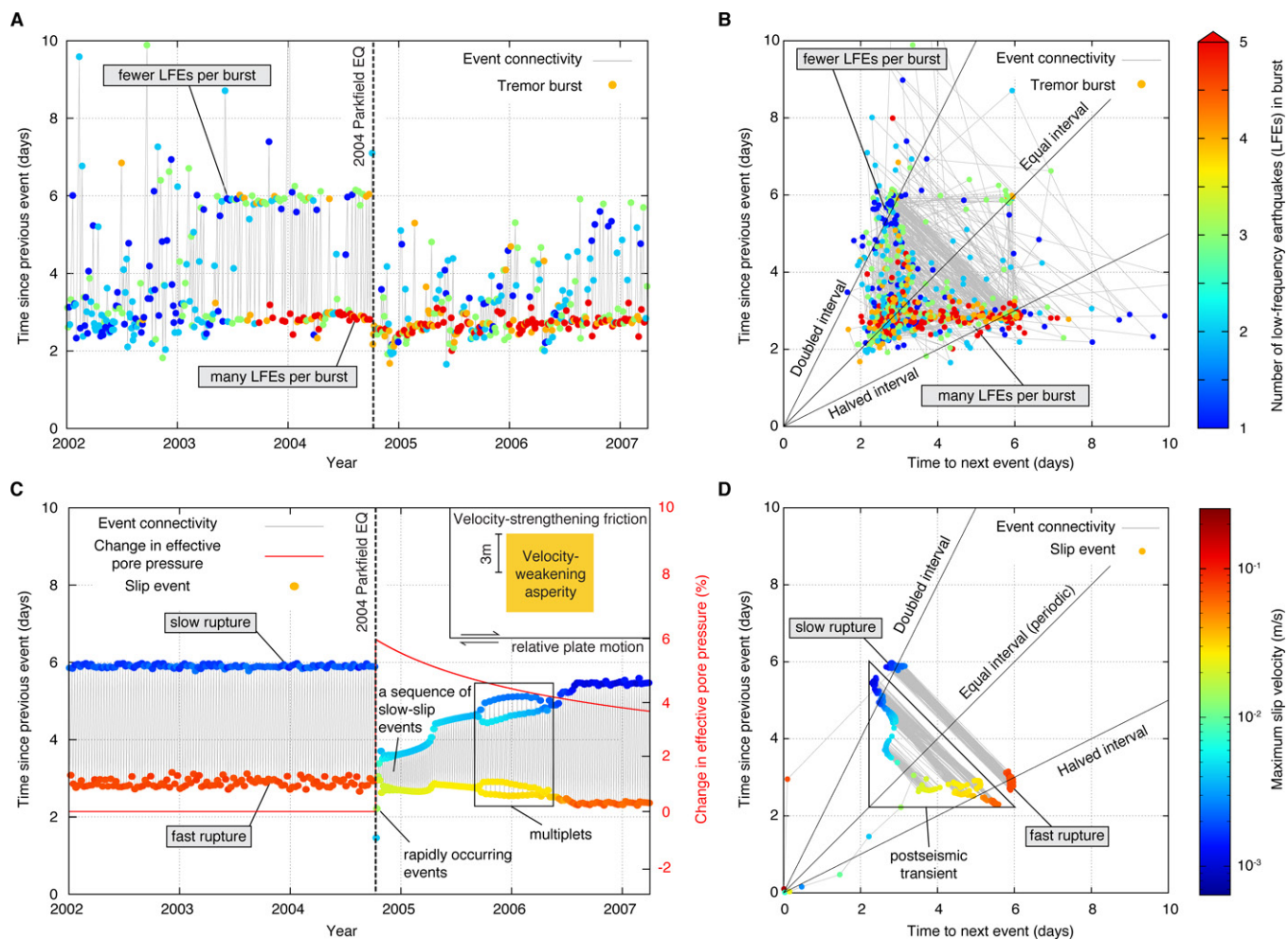
29. Lapusta, N. & Liu, Y. Three-dimensional boundary integral modeling of spontaneous earthquake sequences and aseismic slip. *J. Geophys. Res.* **114**, B09303 (2009).
30. Harris, R. A. *et al.* The SCEC/USGS dynamic earthquake rupture code verification exercise. *Seismol. Res. Lett.* **80**, 119–126 (2009).
31. Chen, T. & Lapusta, N. Scaling of small repeating earthquakes explained by interaction of seismic and aseismic slip in a rate and state fault model. *J. Geophys. Res.* **114**, B01311 (2009).
32. Kaneko, Y., Avouac, J.-P. & Lapusta, N. Towards inferring earthquake patterns from geodetic observations of interseismic coupling. *Nature Geosci.* **3**, 363–369 (2010).
33. Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
34. Ruina, A. Slip instability and state variable friction laws. *J. Geophys. Res.* **88**, 10359–10370 (1983).
35. Rice, J. R. & Ruina, A. L. Stability of steady frictional slipping. *J. Appl. Mech.* **50**, 343–349 (1983).
36. Dieterich, J. H. in *Treatise on Geophysics* Vol. 4, *Earthquake Seismology* (ed. Kanamori, H.) 107–129 (Elsevier Science, 2007).
37. Rice, J. R. & Ben-Zion, Y. Slip complexity in earthquake fault models. *Proc. Natl. Acad. Sci.* **93**, 3811–3818 (1996).
38. Marone, C. & Kilgore, B. Scaling of the critical slip distance for seismic faulting with shear strain in fault zones. *Nature* **362**, 618–621 (1993).
39. Rice, J. R., Lapusta, N. & Ranjith, K. Rate and state dependent friction and the stability of sliding between elastically deformable solids. *J. Mech. Phys. Solids* **49**, 1865–1898 (2001).
40. Beeler, N. M., Thomas, A., Bürgmann, R. & Shelly, D. R. Inferring fault rheology from low-frequency earthquakes on the San Andreas. *J. Geophys. Res.* **118**, 5976–5990 (2013).
41. Peng, Z., Vidale, J. E., Wech, A. G., Nadeau, R. M. & Creager, K. C. Remote triggering of tremor along the San Andreas Fault in central California. *J. Geophys. Res.* **114**, B00A06 (2009).
42. Shelly, D. R., Peng, Z., Hill, D. P. & Aiken, C. Triggered creep as a possible mechanism for delayed dynamic triggering of tremor and earthquakes. *Nature Geosci.* **4**, 384–388 (2011).
43. Thomas, A. M., Bürgmann, R., Shelly, D. R., Beeler, N. M. & Rudolph, M. L. Tidal triggering of low frequency earthquakes near Parkfield, CA: implications for fault mechanics within the brittle-ductile transition. *J. Geophys. Res.* **117**, B05301 (2012).
44. Zigone, D. *et al.* Triggering of tremors and slow slip event in Guerrero, Mexico, by the 2010 Mw 8.8 Maule, Chile, earthquake. *J. Geophys. Res.* **117**, B09304 (2012).
45. Noda, H. & Lapusta, N. Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature* **493**, 518–521 (2013).
46. Barbot, S. & Fialko, Y. A unified continuum representation of post-seismic relaxation mechanisms: semi-analytic models of afterslip, poroelastic rebound and viscoelastic flow. *Geophys. J. Int.* **182**, 1124–1140 (2010).



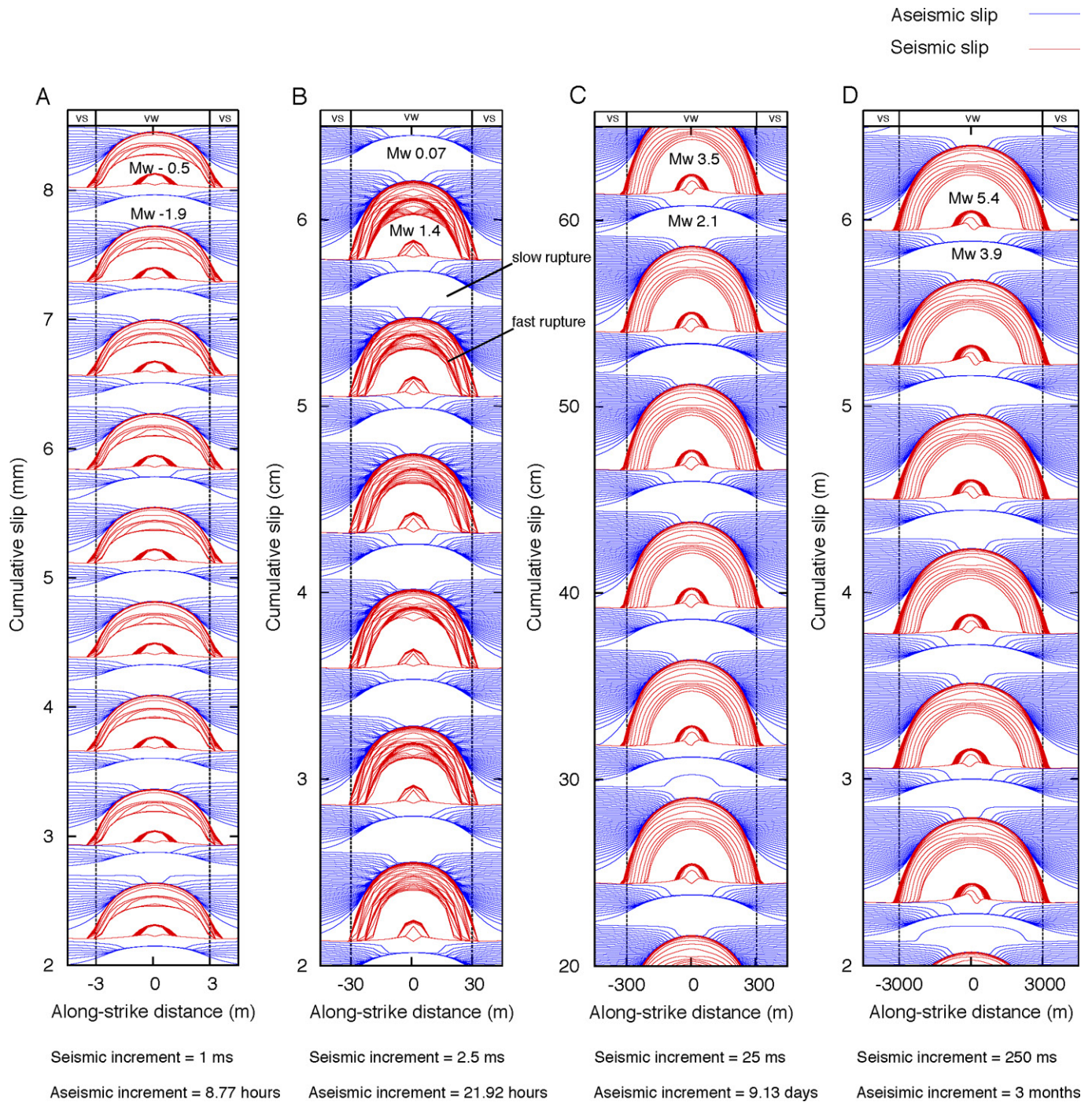
Extended Data Figure 1 | Numerical model of the Parkfield tremors.

a, Schematic diagram of the model geometry and frictional properties. A square velocity-weakening patch (yellow) is embedded in a large velocity-strengthening fault domain. R is the asperity size of the patch. The entire fault domain is driven by a plate rate of $V_{pl} = 3\text{ cm yr}^{-1}$. h^* (dotted circle) is the critical nucleation size of the velocity-weakening patch to generate elasto-dynamic slip. We use the ratio $R/h^* = 0.73$. The model has $R = 3\text{ m}$,

$L = 5\text{ }\mu\text{m}$ and an effective confining pressure of $\bar{\sigma} = 50\text{ MPa}$, but period-doubling recurrence intervals at three and six days can be obtained for various asperity sizes, as discussed in the text. **b**, The yellow square patch in **a** produces a succession of slow and fast ruptures. Fast ruptures may be detected in seismograms and may coincide with one of the LFEs. The model reproduces the period-doubling recurrence pattern of the Parkfield tremors before the 2004 Parkfield earthquake.

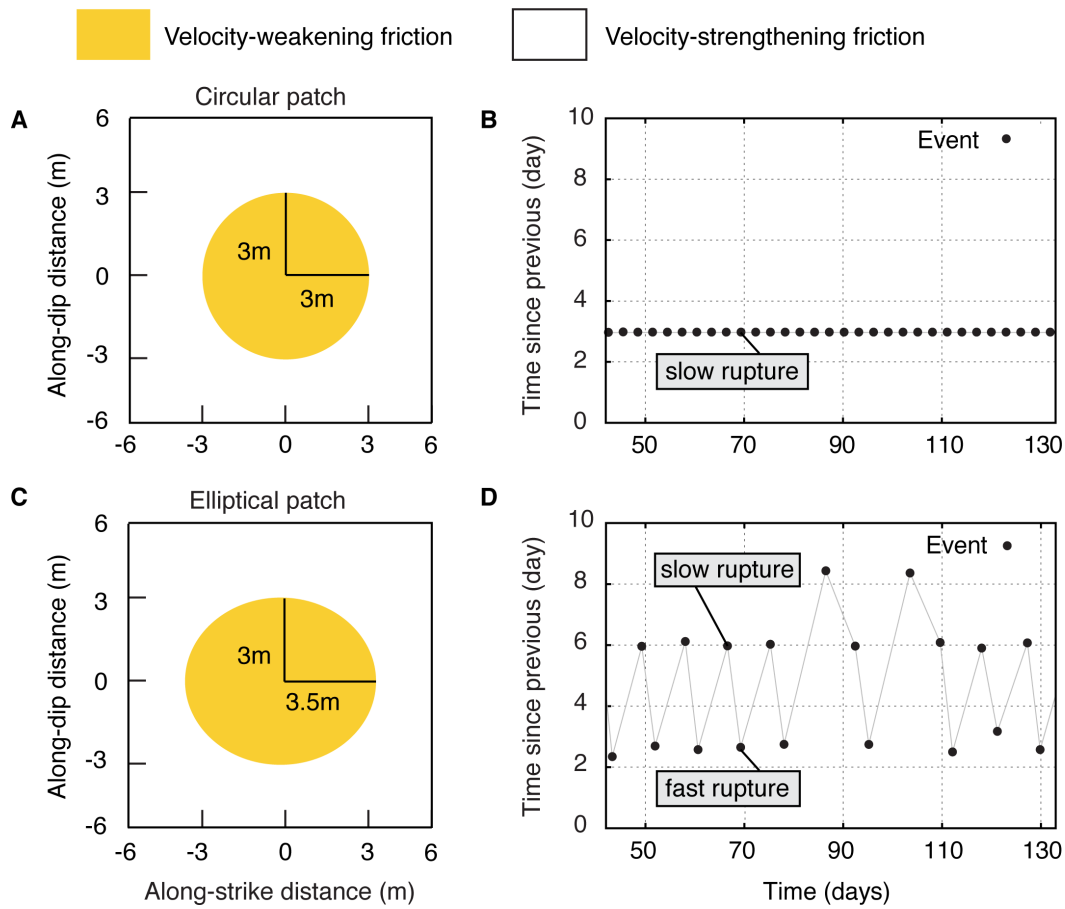


Extended Data Figure 2 | Recurrence pattern and source characteristics of the period-doubling Parkfield tremors and numerical simulations with slip velocity. a–d, Same as Fig. 2, but plotting slip velocity (instead of slip) in c and d.



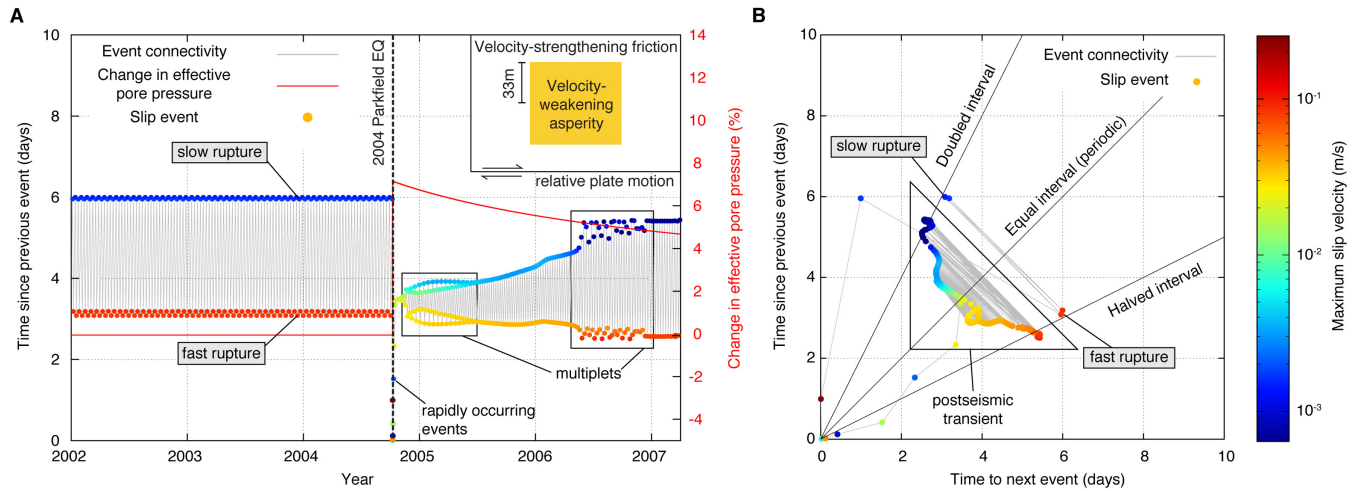
Extended Data Figure 3 | Simulations of period-doubling sequences of slow and fast ruptures along eight orders of magnitude. All simulations assume the same rheological parameters except for the characteristic weakening distance L , which is used to adjust the ratio R/h^* to within the period-doubling range. **a**, With $R = 3$ m and $L = 5 \mu\text{m}$, we obtained slow and fast ruptures with an equivalent geodetic moment magnitude of $M_w -1.9$ and $M_w -0.5$. **b**, With $R = 30$ m and $L = 0.05$ mm, the range is $M_w 0.07$ and $M_w 1.4$. **c**, For $R = 300$ m and $L = 0.5$ mm, the range is

$M_w 2.1$ and $M_w 3.5$. **d**, For $R = 3$ km and $L = 5$ mm, the range is $M_w 3.9$ and $M_w 5.4$. Square patches of velocity-weakening friction with different sizes may exhibit period doubling at almost all magnitudes. The blue and red lines show aseismic and seismic slip, respectively. VS and VW represent the extent of the velocity strengthening and velocity weakening regions, respectively. Increments refer to the time steps between consecutive profiles (larger for aseismic than for seismic slip).



Extended Data Figure 4 | Potential of non-circular shaped asperities to produce complex sequences of slow and fast ruptures with period doubling. **a, c,** Velocity-weakening (VW) patches with two different shapes. **b, d,** The corresponding history of the period separating each event (**a** and **c** respectively). For both simulations, we assume the same frictional parameters and confining pressure. Elliptical (and potentially other non-circular) patches can produce period doubling over a wider range of

parameters than a circular asperity can. With the parameters considered, the non-circular asperities produce period doubling of recurrence time, but the circular patch does not. The parameters considered are 50 MPa, $a = 10^{-2}$, $L = 5 \mu\text{m}$, and the velocity-weakening regions are characterized by $(a - b) = -4 \times 10^{-3}$, while the velocity-strengthening regions have $(a - b) = +4 \times 10^{-3}$. Slip events (seismic and aseismic) are singled out when the slip velocity threshold reaches the value $V_{\text{th}} = 10^{-4} \text{ m s}^{-1}$.



Extended Data Figure 5 | Simulation of the recurrence intervals of the period-doubling Parkfield tremors in a larger slip patch with $R = 33$ m and $\bar{\sigma} = 5$ MPa. **a, Numerical simulation of the Parkfield tremor activity incorporating the change in effective pore pressure (red profile) after the Parkfield earthquake (EQ; vertical dashed line). The coloured dots denote the maximum velocity of the slip events. The slip events are flagged when the slip velocity is above the threshold $V_{pl} = 10^{-4} \text{ m s}^{-1}$ anywhere along the fault. Fast ruptures are preceded by shorter recurrence intervals. A change of pore pressure of about 300 kPa, representing 6% of the effective confining pressure, is applied at the onset of the Parkfield earthquake, resulting in a rapid occurrence of slip events, followed by a**

sequence of multiplets (highlighted in the two boxes) with varying period-multiplying factors. The pressure perturbation is chosen by trial and error to create a two-year transient that is similar to the seismic observations. The top right inset shows the model geometry with an asperity size R of 33 m. **b**, Characterization of the dynamics of the slip events. Fast ruptures are preceded by shorter recurrence intervals. The effective pore-pressure perturbation following the Parkfield earthquake affects the dynamics of the sequence with a temporary interruption of period doubling followed by a gradual recovery. See Fig. 2 legend for details of methodology and presentation.

Revised stratigraphy and chronology for *Homo floresiensis* at Liang Bua in Indonesia

Thomas Sutikna^{1,2*}, Matthew W. Tocheri^{3,4*}, Michael J. Morwood^{1,‡}, E. Wahyu Saptomo^{1,2}, Jatmiko^{1,2}, Rokus Due Awe^{1,2,‡}, Sri Wasisto², Kira E. Westaway⁵, Maxime Aubert^{6,7}, Bo Li¹, Jian-xin Zhao⁸, Michael Storey⁹, Brent V. Alloway^{1,10}, Mike W. Morley¹, Hanneke J. M. Meijer^{4,11}, Gerrit D. van den Bergh¹, Rainer Grün^{12,13}, Anthony Dosseto¹⁴, Adam Brumm^{7,12}, William L. Jungers^{15,16} & Richard G. Roberts¹

Homo floresiensis, a primitive hominin species discovered in Late Pleistocene sediments at Liang Bua (Flores, Indonesia)^{1–3}, has generated wide interest and scientific debate. A major reason this taxon is controversial is because the *H. floresiensis*-bearing deposits, which include associated stone artefacts^{2–4} and remains of other extinct endemic fauna^{5,6}, were dated to between about 95 and 12 thousand calendar years (kyr) ago^{2,3,7}. These ages suggested that *H. floresiensis* survived until long after modern humans reached Australia by ~50 kyr ago^{8–10}. Here we report new stratigraphic and chronological evidence from Liang Bua that does not support the ages inferred previously for the *H. floresiensis* holotype (LB1), ~18 thousand calibrated radiocarbon years before present (kyr cal. BP), or the time of last appearance of this species (about 17 or 13–11 kyr cal. BP)^{1–3,7,11}. Instead, the skeletal remains of *H. floresiensis* and the deposits containing them are dated to between about 100 and 60 kyr ago, whereas stone artefacts attributable to this species range from about 190 to 50 kyr in age. Whether *H. floresiensis* survived after 50 kyr ago—potentially encountering modern humans on Flores or other hominins dispersing through southeast Asia, such as Denisovans^{12,13}—is an open question.

The 2001–2004 excavations at Liang Bua (Fig. 1) revealed the skeletal remains of *H. floresiensis*^{1–3,14–20} at 4–7 m depth in multiple excavated areas (referred to here as Sectors). Charcoal samples collected from similar depths near the eastern wall of the cave (Sector VII) were dated to between about 19 and 13–11 kyr cal. BP^{2,3,7}, whereas an age of ~74 kyr was obtained by coupled electron-spin resonance/uranium-series dating of a *Stegodon florensis insularis* molar recovered from ~4.5 m depth from near the cave centre (Sector IV)^{2,7}. This molar was found ~30 cm below a *H. floresiensis* premolar and ulna (LB2) and ~20 cm and ~130 cm above other postcranial remains (LB10 and LB3, respectively)¹⁸.

The 2007–2014 excavations between and slightly south of these Sectors (Fig. 1c) have revealed new stratigraphic details that are crucial for (re)interpreting the ages inferred originally for *H. floresiensis* (Extended Data Fig. 1 and Supplementary Video 1). The *H. floresiensis*-bearing deposits consist of multiple layers of fine-grained sediment interstratified by layers of weathered limestone, speleothem and loose gravel (Extended Data Fig. 2). These deposits are conformably overlain by an ~2-m-thick sequence of five tephra (referred to here as T1–T5;

Extended Data Fig. 3 and Supplementary Information section 1), separated by clastic sediments and flowstones (Fig. 2). This stratigraphic sequence forms a large pedestal that extends ~12 m laterally from the eastern wall to the cave centre, and is at least 6 m long from north to south. The pedestal is thickest (~4 m) in the middle rear of the cave, where it extends to within ~2 m of the present surface of the cave floor.

The pedestal deposits have been truncated by one or more phases of erosion, resulting in an erosional surface (that is, an unconformity) that slopes steeply down towards the cave mouth (Fig. 3 and Extended Data Figs 4 and 5). The specific timing and nature of the erosional events responsible for this unconformity remain unknown, but presumably include slope wash and other hydrogeological mechanisms, given the active karst setting of the cave. This unconformity was not recognized during the 2001–2004 excavations, raising serious questions about the accuracy of previous age estimates^{2,3,7} for *H. floresiensis*, which were based on inferred stratigraphic associations. To address these questions, we have applied several dating methods to the main stratigraphic units within this pedestal and to the skeletal remains of *H. floresiensis* and other fauna derived from this depositional sequence (see Supplementary Information sections 2–6 for details of each dating method).

Sediment samples for infrared stimulated luminescence (IRSL) and thermoluminescence (TL) dating^{21,22} (Extended Data Figs 6 and 7) were collected from the *H. floresiensis*-bearing deposits directly underlying T1 in the south baulks of Sectors XXI and XXIII, more than 2 m south of the erosional unconformity. These samples gave statistically indistinguishable weighted mean IRSL and TL ages ($\pm 1\sigma$) of 65 ± 5 kyr (LB12-OSL3 to -OSL6) and 71 ± 13 kyr (LB12-23-2 and -4), respectively, for the time since sand-sized grains of feldspar and quartz were last exposed to sunlight. A TL age of 89 ± 7 kyr (LB08-15-3) was obtained for the basal *H. floresiensis*-bearing deposits in Sector XV, while two further samples from the immediately underlying, gravel-rich layer (in Sectors XXI and XXIII) gave IRSL and TL ages of 128 ± 17 kyr (LB12-OSL7) and 113 ± 9 kyr (LB12-23-1), respectively. We also used ²³⁴U/²³⁰Th methods to date four samples of *in situ* speleothem from the *H. floresiensis*-bearing deposits in Sectors XII and XVII (LB07-SXII-F4 and LB09-SXVII-F1 to -F3), which gave ages ($\pm 2\sigma$) of between 80.8 ± 0.8 and 58.3 ± 0.5 kyr.

¹Centre for Archaeological Science, School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. ²Pusat Penelitian Arkeologi Nasional, Jakarta 12510, Indonesia. ³Department of Anthropology, Lakehead University, Thunder Bay, Ontario P7B 5E1, Canada. ⁴Human Origins Program, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. ⁵Traps MQ Luminescence Dating Facility, Department of Environmental Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. ⁶Research Centre for Human Evolution, Place, Evolution and Rock Art Heritage Unit, Griffith University, Gold Coast, Queensland 4222, Australia. ⁷School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. ⁸School of Earth Sciences, University of Queensland, Brisbane, Queensland 4072, Australia. ⁹QUADLAB, Section of Earth and Planetary System Science, Natural History Museum of Denmark, 1350 Copenhagen, Denmark. ¹⁰School of Geography, Environment and Earth Sciences, Victoria University of Wellington, Wellington 6012, New Zealand. ¹¹Department of Natural History, University Museum of Bergen, University of Bergen, 5007 Bergen, Norway. ¹²Research Centre for Human Evolution, Environmental Futures Research Institute, Griffith University, Brisbane, Queensland 4111, Australia. ¹³Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory 0200, Australia. ¹⁴GeoQUEST Research Centre, School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. ¹⁵Department of Anatomical Sciences, Stony Brook University Medical Center, Stony Brook, New York 11794, USA. ¹⁶Association Vahatra, BP 3972, Antananarivo 101, Madagascar.

*These authors contributed equally to this work.

‡Deceased.

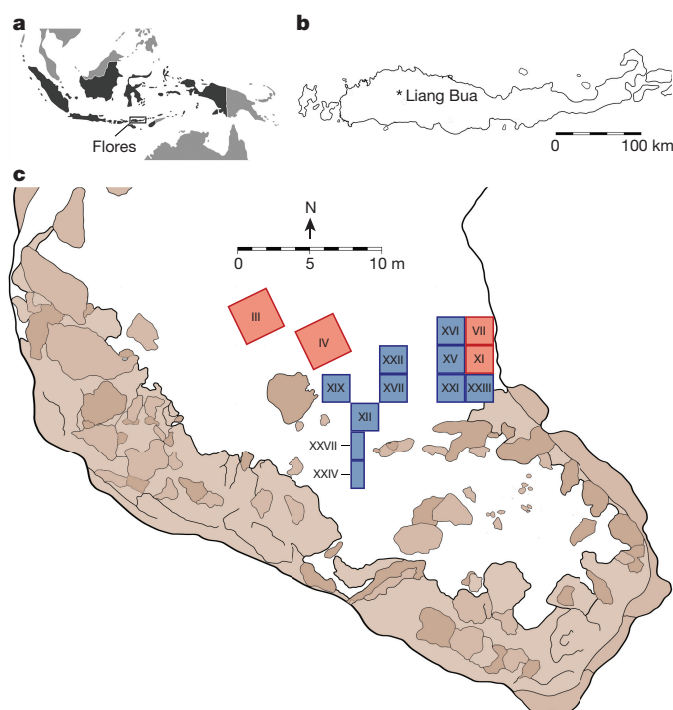


Figure 1 | Site location. **a**, Location of Flores within Indonesia. **b**, Location of Liang Bua on Flores. **c**, Site plan of Sectors discussed in the text (the 2001–2004 and 2007–2014 excavations are shaded red and blue, respectively). The remaining cave floor sediments are shaded white, while the areas shaded brown are exposed rocks, stalagmites and other surfaces covered in speleothems.

Three *H. floresiensis* ulnae and seven *Stegodon* bones were sampled for $^{234}\text{U}/^{230}\text{Th}$ dating, as well as a modern human femoral shaft fragment from the Holocene deposits in Sector IV (Extended Data Figs 8 and 9). Owing to the open system behaviour of bone, such analyses usually yield minimum ages for bone deposition and subsequent uranium uptake²³. The modelled age²⁴ of ~ 7.5 kyr for the modern human bone (132A/LB/IV/27D/03) is consistent with charcoal ^{14}C ages for the underlying and overlying sediments (~ 9.5 and 6.4 kyr cal. BP, respectively⁷). The *H. floresiensis* ulnae have modelled $^{234}\text{U}/^{230}\text{Th}$ ages ($\pm 2\sigma$) for individual laser-ablation tracks that range from 86.9 ± 7.9 to 71.5 ± 4.3 kyr for LB1 (Sectors VII and XI), 71.4 ± 1.1 to 66.7 ± 0.8 kyr for LB2 (Sector IV), and 66.0 ± 4.3 to 54.6 ± 2.1 kyr for LB6 (Sector XI). The *Stegodon* bone samples (all from Sector XI) span a modelled age range of 80.6 ± 11.3 to 40.5 ± 2.0 kyr, with the youngest minimum age deriving from a bone (U-s-05/LB/XI/51/04) recovered from the same sediments and depth as LB6. Delayed diffusion of uranium into the dense bone matrix of *Stegodon* may account for the youngest minimum ages appearing more recent than those obtained for the bones of *H. floresiensis*, the speleothems and the sediments.

We dated T1, which directly overlies the *H. floresiensis*-bearing deposits, using $^{40}\text{Ar}/^{39}\text{Ar}$ methods^{25,26}. The inverse isochron age ($\pm 1\sigma$) of 79 ± 12 kyr (Extended Data Fig. 3) is imprecise owing to the low yield of radiogenic argon from the hornblende crystals, but the 2σ age range (103 – 55 kyr) is consistent with the luminescence and $^{234}\text{U}/^{230}\text{Th}$ ages for the underlying samples. Immediately overlying T1 is a series of interstratified sedimentary units, including T2 and several flowstones. These are in turn overlain by T3, a ~ 0.75 m-thick volcanoclastic mass flow deposit, formerly referred to as the ‘black tuff’ or similar^{18,27}.

Sediments between T2 and T3 in Sector XXI yielded a weighted mean IRSL age of 66 ± 9 kyr (LB12-OSL1 and -OSL2) and a TL age of 59 ± 13 kyr (LB12-23-3). Interbedded flowstones and a small stalagmite in the same Sector gave $^{234}\text{U}/^{230}\text{Th}$ ages of between 66.1 ± 0.3 and 54.4 ± 0.3 kyr (LB-S.XXI 5 #08, LB-S.XXI 8-T #09 and 8-B #10, and LB/S.XXI 10-01 and -01R), while in Sector XII—nearer the cave

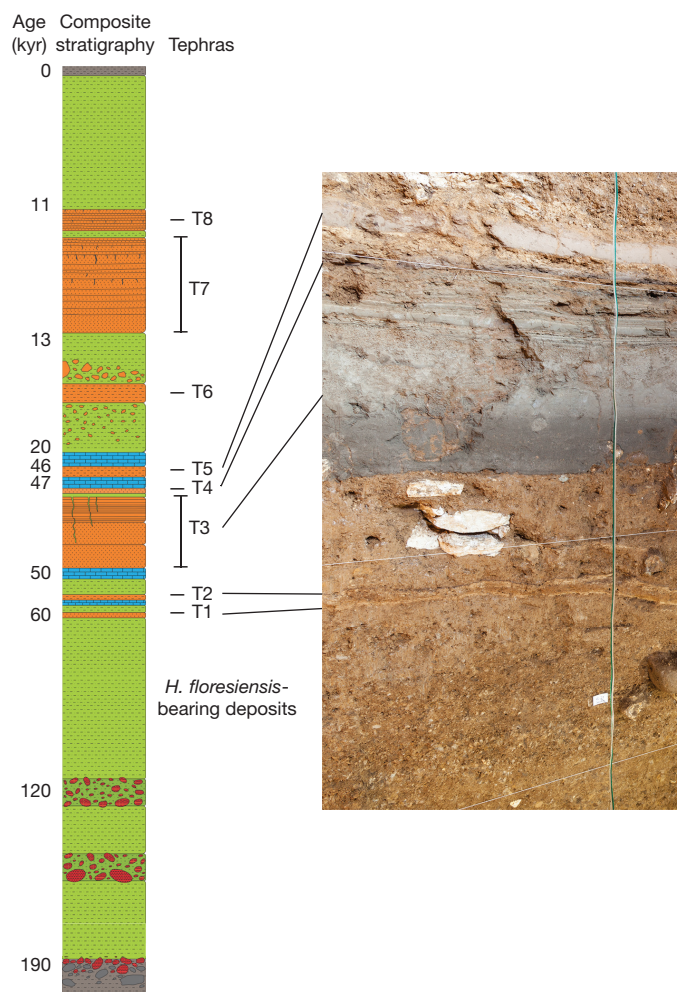


Figure 2 | Composite stratigraphic section of deposits at Liang Bua, with approximate ages. The deposits accumulated above a fluvial conglomerate and are capped by recent sediments (both shown in grey). Skeletal remains of *Homo floresiensis* occur in deposits stratigraphically beneath a sequence of eight volcanic tephra (T1–T8), separated by calcitic speleothems (blue) and fine-grained clastic sediments (green). As the thickness, grain size and slope angle of each unit vary considerably within the cave, only the approximate relative thicknesses of the units discussed in the text are shown; the minimum depth of this composite section would exceed 15 m. See Fig. 3 and Extended Data Fig. 5a for three-dimensional representations of the stratigraphy. Also indicated are units with concentrations of rounded, gravel-size clasts of igneous rock (red) or irregularly shaped, eroded fragments of T1, T2 and T3 (orange), and units with signs of bioturbation (upper parts of T3 and T7). The photograph shows T1–T5 and interstratified sediments that conformably overlie the *H. floresiensis*-bearing deposits (south baulk of Sector XXI).

centre—a flowstone immediately underlying T3 (LB07-SXII-F1) was dated to 49.6 ± 0.5 kyr. Hominin skeletal remains have not been recovered from the deposits between T2 and T3. However, stone artefacts are present and these display similar raw material proportions to those found underneath T1 ($\sim 80\%$ silicified tuff and $\sim 20\%$ chert), in contrast to the preference of modern humans for chert ($\sim 60\%$) observed at Liang Bua during the Holocene⁴. Also present in these sediments and those underlying T1 are the remains of other extinct large-bodied taxa, including *Stegodon*, giant marabou stork and vulture^{5,6}. The specific processes that ultimately resulted in the extinction of these taxa and *H. floresiensis* remain poorly understood.

Overlying T3 is T4, which is covered by loose, coarse gravel in two of the southernmost Sectors excavated (XII and XXVII). Three flowstone samples^{7,28} (weighted mean $^{234}\text{U}/^{230}\text{Th}$ age of 46.6 ± 0.5 kyr) directly overlie the southern parts of this coarse gravel layer. From the eastern

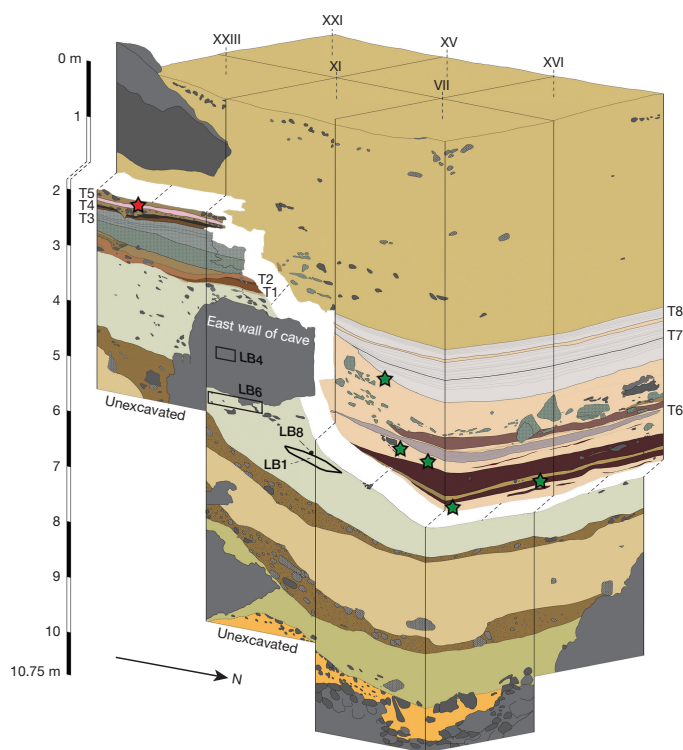


Figure 3 | Stratigraphy of excavated Sectors near the eastern wall of the cave. Multiple specimens of *Homo floresiensis* (LB1, LB4, LB6 and LB8) were recovered previously^{1–3,14–20} from sediments now recognized to directly underlie a sequence of five tephtras (T1–T5) and interstratified deposits. Together, these remnant deposits form a pedestal, the top of which is dated to ~46 kyr cal. BP (charcoal sample denoted by the red star at ~2 m depth in Sector XXIII). The overlying section—separated here by a white band and dotted lines for emphasis—represents deposits (including three additional tephtras, T6–T8) that rest unconformably on the steeply sloping erosional surface of the pedestal. The green stars in Sector VII mark the respective locations (from top to bottom) of the charcoal samples dated to approximately 13.0, 18.5, 18.1, 19.0 and 19.2 kyr cal. BP, which were used erroneously^{2,3,7} to infer the latest occurrences of *H. floresiensis*.

wall to the cave centre, T4 is overlain by two flowstone layers that are separated from each other by T5. At 1.89 m depth in Sector XXIII, charcoal (sample OxA-X-2648-13) from immediately above the upper surface of T5 yielded a ¹⁴C age of ~46 kyr cal. BP (95% confidence interval of 47.7–44.1 kyr cal. BP).

In contrast, all of the charcoal samples used previously to infer ages of ~18 kyr cal. BP for LB1 and about 17 or 13–11 kyr cal. BP as the *terminus ante quem* for *H. floresiensis* and *Stegodon*^{2,3,5,7} originate from deposits that unconformably overlie the remnant pedestal, based on their plotted coordinates (Fig. 3). These younger sediments are interstratified by three additional tephtras (T6–T8; Extended Data Fig. 10). Tephra T6 is observed at ~5.8 m depth in the northern parts of Sectors VII and XVI, followed by T7 and T8 (formerly referred to as the ‘white’ tuffaceous silts^{2,3,18,27}) at about 3.5–5 m depth (Sectors III, IV, VII, XI, XV, XVI and XXII). The onset of accumulation of the depositional sequence above the unconformity ~20 kyr ago is consistent with a TL age of 23 ± 7 kyr for sediments collected less than 15 cm above the unconformity in Sector XVI (LB09-16-2) and two new ¹⁴C ages of ~12.7 kyr cal. BP for charcoal recovered from just beneath T7 in Sector XXII (D-AMS 005953 and 005954).

Portions of the pedestal have been reworked into this younger depositional sequence. In Sector XXII, a displaced slab of intact pedestal deposit rests above the unconformity but underlies T7 (Extended Data Fig. 10c, d). In other Sectors, eroded fragments of T1, T2 and T3 are consistently observed in the younger, overlying deposits^{2,3,18}.

We investigated the possibility that the luminescence ages for the sediment samples collected in 2003 from alongside LB1 and ~1 m above it² were compromised by inadvertently sampling across the unconformity and measuring multiple grains simultaneously from this mixture. The wide spread in ages (from about 170 to 10 kyr ago) obtained for individual feldspar grains using newly developed IRSL dating procedures²¹ supports this suggestion (Extended Data Fig. 7k, l).

The new stratigraphic and chronological evidence for Liang Bua indicates that a pedestal of remnant deposits, dating to more than ~46 kyr cal. BP, has an erosional upper surface that slopes steeply downwards to the north and is unconformably overlain by sediments younger than ~20 kyr cal. BP. All skeletal remains assigned to *H. floresiensis* are from the pedestal deposits dated to approximately 100–60 kyr ago, while stone artefacts reasonably attributable to this species range from about 190 kyr²⁸ to 50 kyr in age. Parts of southeast Asia may have been inhabited by Denisovans^{12,13} or other hominins^{29,30} during this period, and modern humans had reached Australia by 50 kyr ago^{8–10}. But whether *H. floresiensis* survived after this time, or encountered modern humans, Denisovans or other hominin species on Flores or elsewhere, remain open questions that future discoveries may help to answer.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 August 2015; accepted 28 January 2016.

Published online 30 March 2016.

- Brown, P. *et al.* A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
- Morwood, M. J. *et al.* Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* **431**, 1087–1091 (2004).
- Morwood, M. J. *et al.* Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* **437**, 1012–1017 (2005).
- Moore, M. W., Sutikna, T., Jatmiko, Morwood, M. J. & Brumm, A. Continuities in stone flaking technology at Liang Bua, Flores, Indonesia. *J. Hum. Evol.* **57**, 503–526 (2009).
- van den Bergh, G. D. *et al.* The Liang Bua faunal remains: a 95 kyr sequence from Flores, East Indonesia. *J. Hum. Evol.* **57**, 527–537 (2009).
- Meijer, H. J. M. *et al.* Late Pleistocene–Holocene non-passerine avifauna of Liang Bua (Flores, Indonesia). *J. Vertebr. Paleontol.* **33**, 877–894 (2013).
- Roberts, R. G. *et al.* Geochronology of cave deposits at Liang Bua and of adjacent river terraces in the Wae Racang valley, western Flores, Indonesia: a synthesis of age estimates for the type locality of *Homo floresiensis*. *J. Hum. Evol.* **57**, 484–502 (2009).
- Roberts, R. G. *et al.* Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia. *Nature* **345**, 153–156 (1990).
- Bowler, J. M. *et al.* New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* **421**, 837–840 (2003).
- Clarkson, C. *et al.* The archaeology, chronology and stratigraphy of Madjedbebe (Malakunanja II): a site in northern Australia with early occupation. *J. Hum. Evol.* **83**, 46–64 (2015).
- Morwood, M. J. *et al.* Preface: research at Liang Bua, Flores, Indonesia. *J. Hum. Evol.* **57**, 437–449 (2009).
- Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
- Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- Falk, D. *et al.* The brain of *Homo floresiensis*. *Science* **308**, 242–245 (2005).
- Larson, S. G. *et al.* *Homo floresiensis* and the evolution of the hominin shoulder. *J. Hum. Evol.* **53**, 718–731 (2007).
- Tocheri, M. W. *et al.* The primitive wrist of *Homo floresiensis* and its implications for hominin evolution. *Science* **317**, 1743–1745 (2007).
- Jungers, W. L. *et al.* The foot of *Homo floresiensis*. *Nature* **459**, 81–84 (2009).
- Morwood, M. J. & Jungers, W. L. (Eds). Paleanthropological Research at Liang Bua, Flores, Indonesia. *J. Hum. Evol.* **57**, 437–648 (2009).
- Kaifu, Y. *et al.* Craniofacial morphology of *Homo floresiensis*: description, taxonomic affinities, and evolutionary implication. *J. Hum. Evol.* **61**, 644–682 (2011).
- Orr, C. M. *et al.* New wrist bones of *Homo floresiensis* from Liang Bua (Flores, Indonesia). *J. Hum. Evol.* **64**, 109–129 (2013).
- Li, B., Jacobs, Z., Roberts, R. G. & Li, S.-H. Review and assessment of the potential of post-IR IRSL dating methods to circumvent the problem of anomalous fading in feldspar luminescence. *Geochronometria* **41**, 178–201 (2014).
- Westaway, K. E. & Roberts, R. G. A dual-aliquot regenerative-dose protocol (DAP) for thermoluminescence (TL) dating of quartz sediments using the light-sensitive and isothermally stimulated red emissions. *Quat. Sci. Rev.* **25**, 2513–2528 (2006).

23. Grün, R., Eggins, S., Kinsley, L., Moseley, H. & Sambridge, M. Laser ablation U-series analysis of fossil bones and teeth. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **416**, 150–167 (2014).
24. Sambridge, M., Grün, R. & Eggins, S. U-series dating of bone in an open system: the diffusion-adsorption-decay model. *Quat. Geochronol.* **9**, 42–53 (2012).
25. Storey, M., Roberts, R. G. & Saidin, M. Astronomically calibrated $^{40}\text{Ar}/^{39}\text{Ar}$ age for the Toba supereruption and global synchronization of late Quaternary records. *Proc. Natl Acad. Sci. USA* **109**, 18684–18688 (2012).
26. Rivera, T. A., Storey, M., Schmitz, M. D. & Crowley, J. L. Age intercalibration of $^{40}\text{Ar}/^{39}\text{Ar}$ sanidine and chemically distinct U/Pb zircon populations from the Alder Creek Rhyolite Quaternary geochronology standard. *Chem. Geol.* **345**, 87–98 (2013).
27. Westaway, K. E. *et al.* *Homo floresiensis* and the late Pleistocene environments of eastern Indonesia: defining the nature of the relationship. *Quat. Sci. Rev.* **28**, 2897–2912 (2009).
28. Westaway, K. E. *et al.* Establishing the time of initial human occupation of Liang Bua, western Flores, Indonesia. *Quat. Geochronol.* **2**, 337–343 (2007).
29. Mijares, A. S. *et al.* New evidence for a 67,000-year-old human presence at Callao Cave, Luzon, Philippines. *J. Hum. Evol.* **59**, 123–132 (2010).
30. van den Bergh, G. D. *et al.* Earliest hominin occupation of Sulawesi, Indonesia. *Nature* **529**, 208–211 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements The 2007–2014 excavations at Liang Bua were supported by an Australian Research Council (ARC) Discovery Project grant to M.J.M. (DP0770234), a Waitt Foundation/National Geographic Society grant to M.W.T. and T.S. (No. 2121-2) and a Smithsonian Scholarly Studies Program award to M.W.T. Additional funding was provided by the Peter Buck Fund for Human Origins Research, the Smithsonian's Human Origins Program, the University of

Wollongong (UOW) and the ARC (DP1093049 to K.E.W.). T.S. is supported by a UOW postgraduate scholarship, M.W.T. by a Canada Research Chair, M.A. and A.B. by ARC Discovery Early Career Researcher Awards (DE140100254 and DE130101560, respectively), B.L. by an ARC Future Fellowship (FT14010038), R.G.R. by an ARC Australian Laureate Fellowship (FL130100116) and B.V.A. by a Victoria University of Wellington Science Faculty Research Grant (201255). QUADLAB is funded by the Villum Foundation. Fieldwork was authorised by Pusat Penelitian Arkeologi Nasional (Jakarta, Indonesia) and Pemerintah Daerah Kabupaten Manggarai (Flores, Nusa Tenggara Timur). We also thank I Made Geria, V. N. Sene, R. Potts, P. Goldberg, K. Douka, G. Veatch, V. Rossi, A. Metallo, L. Kinsley, Y. Jafari, T. Lachlan, A. D. Nguyen, D. Yurnaldi, R. Setiawan, I Dewa Kompiang and the entire Liang Bua Team from Teras, Golo Manuk and Bere.

Author Contributions M.J.M., R. P. Soejono and R.G.R. conceived and coordinated the original research program at Liang Bua (2001–2004). The new excavations were planned and directed by T.S., E.W.S. and M.J.M. (2007–2009), and by T.S., M.W.T., E.W.S., J. and M.J.M. (2010–2014). T.S. led the stratigraphic analyses, with major contributions from M.W.T., S.W., M.J.M., K.E.W., R.D.A., E.W.S. and J., and additional input from M.W.M., H.J.M.M., G.D.vdB., B.V.A., A.B., W.L.J. and R.G.R. Dating analyses were conducted by B.L. and R.G.R. (IRSL), K.E.W. (TL), M.A., R.G. and A.D. ($^{234}\text{U}/^{230}\text{Th}$, bones), J.-x.Z. ($^{234}\text{U}/^{230}\text{Th}$, speleothems), and M.S. ($^{40}\text{Ar}/^{39}\text{Ar}$). B.V.A. analysed the volcanic tephra, R.D.A., H.J.M.M., G.D.vdB., M.W.T. and W.L.J. analysed the faunal remains, and J. analysed the stone artefacts. T.S., M.W.T. and R.G.R. wrote the paper, with early contributions from M.J.M. and additional input from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.S. (thomasutikna@gmail.com), M.W.T. (tocherim@gmail.com) and R.G.R. (rgrob@uow.edu.au).

METHODS

Archaeological excavation. The 2007–2014 excavations at Liang Bua, including eight 2×2 m and two 1×2 m areas (referred to as Sectors), proceeded in 10 cm intervals (referred to as spits) while following stratigraphic units. Timber shoring of the baulks was installed after ~ 2.5 m depth for safety. *In situ* findings (for example, bones, artefacts and charcoal) were plotted in three dimensions and the sediments from each spit were sieved by hand, followed by wet sieving (2 mm mesh). All recovered findings were cleaned, catalogued and transported to Pusat Penelitian Arkeologi Nasional (Jakarta, Indonesia) for curation and further study.

Electron microprobe analysis of volcanic glass. Bulk samples of four tephra were wet sieved at $>32 \mu\text{m}$ and then dry sieved into $32\text{--}63 \mu\text{m}$, $63\text{--}125 \mu\text{m}$, $125\text{--}250 \mu\text{m}$, $250\text{--}500 \mu\text{m}$ and $500 \mu\text{m}\text{--}1\text{ mm}$ fractions. Depending on the grain-size distribution of the sample, the dominant dry fraction containing the most glass material was mounted into an epoxy resin. Glass from bubble-wall shards or vesiculated pumice fragments was analysed as individual grains ($>63 \mu\text{m}$) using an electron microprobe. Major element determinations were made using a JEOL Superprobe (JXA-8230) housed at Victoria University of Wellington, New Zealand, using the ZAF correction method³¹. Analyses were performed with 15 kV accelerating voltage, 8 nA beam current, and an electron beam defocused to $10\text{--}20 \mu\text{m}$. Standardization was achieved by means of mineral and glass standards. Rhyolitic glass standard ATHO-G (ref. 32) was routinely used to monitor calibration in all analytical runs and to evaluate any day-to-day differences in calibration. All analyses were normalized to 100% (by weight) anhydrous and total iron was calculated as FeO (Supplementary Information section 1). The major element compositions of the four tephra and the Youngest Toba Tuff³³ are displayed as bivariate plots in Extended Data Fig. 3c–e.

Infrared stimulated luminescence (IRSL) dating of feldspar grains. Sediment samples were collected using opaque plastic tubes hammered horizontally into cleaned stratigraphic sections and wrapped in black plastic after removal. Field measurements of the gamma dose rate at each sample location were made using a portable gamma-ray detector (Exploranium GR-320) inserted into the emptied tube holes. Sediment samples were also collected from the tube holes for laboratory determinations of water content and beta dose rate at the University of Wollongong. The environmental dose rate of each sample was calculated as the sum of the beta dose rate (estimated from beta-counting of dried and powdered sediment samples using a Riso low-level beta multicounter system and allowing for beta-dose attenuation and other factors³⁴), the *in situ* gamma dose rate, and the estimated cosmic-ray dose rate³⁵. The latter took account of the burial depth of each sample, the thickness of cave roof overhead, the zenith-angle dependence of cosmic rays, and the latitude, longitude and altitude of Liang Bua. Each of these external dose rate contributors were adjusted for long-term water content. The measured (field) water contents of the seven samples collected in 2012 range from 26 to 35%, but higher and lower field water contents have been reported for other sediment samples collected at Liang Bua: 15–30% (ref. 2), 3–22% (ref. 7) and 24–39% (Supplementary Information section 2). We used a value of $20 \pm 5\%$ as a mid-range estimate of the long-term water content, with the standard error sufficient to cover (at 2σ) most of the field measurements. The total dose rate of each sample also includes an estimate of the internal beta dose rate due to the radioactive decay of ^{40}K and ^{87}Rb . This estimate was based on a K concentration of $12 \pm 1\%$, determined from energy- and wavelength-dispersive X-ray spectroscopic measurements of individual feldspar grains^{36,37} (see Supplementary Information section 2), and an assumed Rb concentration of $400 \pm 100 \mu\text{g g}^{-1}$ (ref. 38). Potassium-rich feldspar (K-feldspar) grains of $90\text{--}212$ or $180\text{--}212 \mu\text{m}$ diameter were extracted from the sediment samples under dim red illumination and prepared using standard procedures³⁹, including the use of a sodium polytungstate solution of 2.58 g cm^{-3} density to separate the feldspar grains from heavier minerals. The separated grains were also washed in 10% hydrofluoric acid for 40 min to clean their surfaces and reduce the thickness of the alpha-irradiated outer layer by $\sim 15 \mu\text{m}$, which was taken into account in the dose rate calculation. For each sample, single aliquots composed of a few hundred grains were measured using an automated Riso TL/OSL reader equipped with infrared (875 nm) light-emitted diodes (LEDs) for stimulation⁴⁰ and a calibrated $^{90}\text{Sr}/^{90}\text{Y}$ source for beta irradiations. We also made measurements of individual K-feldspar grains ($180\text{--}212 \mu\text{m}$ diameter) from two samples collected in 2012 (LB12-OSL3 and -OSL4) and the two original sediment samples (LBS7-40a and LBS7-42a) collected from above and alongside the remains of LB1 (ref. 2). Infrared stimulation of individual K-feldspar grains was achieved using a focused laser beam (830 nm)⁴⁰. The IRSL emissions from the single aliquots and single grains were detected using an Electron Tubes Ltd 9235B photomultiplier tube fitted with Schott BG-39 and Corning 7-59 filters to transmit wavelengths of $320\text{--}480 \text{ nm}$. To estimate the equivalent dose (D_e) for each aliquot or grain^{39,41}, we initially used the multiple elevated temperature post-infrared IRSL (MET-pIRIR) regenerative-dose procedure^{21,30,42}, in which the IRSL signals are measured by

progressively increasing the stimulation temperature from 50°C to 250°C in 50°C increments. The samples yielded very dim signals, however, so we used a two-step pIRIR procedure^{21,43} to improve the signal-to-noise ratio. Grains were preheated at 320°C for 60 s before infrared stimulation of the natural, regenerative and test doses at 50°C for 200 s. The pIRIR signal was then measured at 290°C for either 200 s (single aliquots) or 1 s (single grains), followed by an infrared bleach at 325°C for 100 s at the end of each regenerative-dose cycle. Example pIRIR decay and dose–response curves are shown in Extended Data Fig. 7a–c. Performance tests of the regenerative-dose procedure and details of residual dose measurements^{21,44} and anomalous fading tests^{45,46} are described in Supplementary Information section 2. We estimated the pIRIR age of each sample from the weighted mean D_e (calculated using the central age model^{47,48}) divided by the environmental dose rate, and applied corrections for residual dose and anomalous fading. For the latter corrections, we used the weighted mean fading rate ($0.9 \pm 0.3\%$ per decade) measured for five of the seven samples collected in 2012.

Thermoluminescence (TL) dating of quartz grains. Sediment samples were taken in the same manner as the IRSL samples, as were the *in situ* gamma dose rate measurements (made using an ORTEC digi-DART gamma-ray detector). Additional sediment samples were also collected for determinations of water content and beta dose rate at Macquarie University. The environmental dose rate of each sample consists of three external components, each adjusted for long-term water content ($20 \pm 5\%$): the beta dose rate (estimated from beta-counting and allowing for beta-dose attenuation), the *in situ* gamma dose rate, and the estimated cosmic-ray dose rate. An assumed internal dose rate of $0.03 \pm 0.01 \text{ Gy kyr}^{-1}$ was also included in the total dose rate. Quartz grains were separated from the sediment samples under dim red illumination and prepared using standard procedures³⁹, including mineral separations using 2.70 and 2.62 g cm^{-3} density solutions of sodium polytungstate to isolate the quartz and a 40% hydrofluoric acid etch for 45 min to remove the alpha-irradiated outer $\sim 20 \mu\text{m}$ of each grain. Aliquots composed of about 10,000 quartz grains were measured using a dual-aliquot regenerative-dose TL protocol²² to isolate the light-sensitive red emissions⁴⁹. This procedure was originally developed for application at Liang Bua² and requires two aliquots of each sample: one to estimate the D_e associated with the heat-reset TL traps and the other to measure the total TL signal, from which the D_e associated with the light-sensitive TL traps is estimated by subtraction. Measurements were made on a Riso TL/OSL reader fitted with an Electron Tubes Ltd 9658B photomultiplier tube and Schott BG-39 and Kopp 2-63 filters designed to transmit red emissions (peak transmission in the $600\text{--}620 \text{ nm}$ range, with minimum and maximum wavelengths of 580 and 670 nm), and cooled to -22°C to reduce the background count rate. Bleaching was performed using a halogen lamp and light guide integrated into the reader, and laboratory doses were given using a $^{90}\text{Sr}/^{90}\text{Y}$ source mounted on the reader. The quartz grains were first heated to 260°C at a heating rate of 5 K s^{-1} and then held at this temperature for 1,000 s to induce an isothermal TL signal and reduce the unwanted glow from incandescence. Following ref. 22, D_e values were estimated from the 20–30 s interval of isothermal TL (which was shown to be light-sensitive) and the final 160 s was used as background. Two of the samples contained abundant quantities of quartz, so problems associated with inter-aliquot variability^{50,51} were reduced by measuring 12 replicates of each pair of aliquots (Supplementary Information section 2). As prolonged unlight exposure is required the empty the light-sensitive TL traps, the ages obtained should be regarded as maximum estimates of the time elapsed since sediment deposition²².

Uranium-series ($^{234}\text{U}/^{230}\text{Th}$) dating of bones. Samples of bone from three specimens of *Homo floresiensis* (LB1, LB2 and LB6), one *Homo sapiens* and eight *Stegodon florensis insularis* were analysed using laser ablation uranium-series isotope measurement analysis procedures and instruments similar to those described previously^{23,52}. Cuts were made perpendicular to the bone surface using a rotatory tool equipped with a thin ($100 \mu\text{m}$ wide) diamond saw blade. The cut samples were mounted into aluminium cups, aligning the cross-sectioned surfaces with the outer rim of the sample holder to position the samples on the focal plane of the laser in the sampling cell. Sequential laser spot analyses were undertaken along 1–5 parallel tracks per sample, starting from the interior of each cross-sectioned bone (Extended Data Figs 8 and 9 and Supplementary Information section 3). Uranium and thorium isotopes were measured at the Australian National University using a Finnigan MAT Neptune multi-collector inductively-coupled plasma mass spectrometer (MC-ICP-MS) equipped with multiple Faraday cups and ion counters. Two ion counters were set to masses of 230.1 and 234.1, while the Faraday cups measured the masses 232, 235 and 238. Samples were ablated using a Lambda Physik LPFPro ArF excimer laser (193 nm) coupled to the MC-ICP-MS through a custom-designed Helex ablation cell. The analyses were made at regular spacing (typically 2–3 mm) along each track, using a laser spot size of $265 \mu\text{m}$ and a 5 Hz pulse rate. The samples were initially cleaned for 5 s and ablation pits were measured for 50 s. These measurements were bracketed by analyses of reference

standards to correct for instrument drift. Semi-quantitative estimates of uranium and thorium concentrations were made from repeated measurements of the SRM NIST-610 glass standard (U: $461.5 \mu\text{g g}^{-1}$, Th: $457.2 \mu\text{g g}^{-1}$) and uranium-isotope (activity) ratios from repeated measurements of rhinoceros tooth dentine from Hexian sample 1118 (ref. 53). Apparent $^{234}\text{U}/^{230}\text{Th}$ ages were estimated for each track using a diffusion-adsorption-decay (DAD) model²⁴, which uses the entire set of isotope measurements made along the track to calculate the rate of diffusion of ^{238}U and ^{234}U into the bone. Bones behave as geochemically 'open' systems, so the diffusion of uranium into a bone may have occurred soon after it was deposited or much later. The original isotopic signature may also be overprinted by secondary uranium uptake that is hard to recognize. As such, uranium-series ages for bones are most likely to be minimum estimates of time since deposition, with the extent of age underestimation being very difficult or impossible to evaluate²³. The modelled ages were calculated after rejecting data points associated with detrital contamination (U/Th elemental ratios of ≤ 300) and data points at the surface of the bones where secondary overprinting was suspected. Results are tabulated in Supplementary Information section 3, with all errors given at 2σ .

Uranium-series ($^{234}\text{U}/^{230}\text{Th}$) dating of speleothems. Samples of calcite deposited as speleothems (flowstones and a 1 cm-tall stalagmite) were collected during excavation. Unlike bone, speleothems commonly behave as geochemically 'closed' systems, with no loss or gain of uranium after calcite crystallization⁵⁴. The cleanest portion of each sample was selected, ground to the size of a rice grain, cleaned ultrasonically and then handpicked, avoiding pieces that appeared to be porous or contaminated. Age determinations were made at the University of Queensland using a Nu Plasma MC-ICP-MS. Uranium and thorium separation procedures, MC-ICP-MS analytical protocols, details of spike calibrations, blank and 'memory' assessments, and repeat measurements of standards have been described previously^{55,56}. The $^{234}\text{U}/^{230}\text{Th}$ ages were calculated using Isoplot 3.75 (ref. 57) and half-lives of 245.25 kyr (^{234}U) and 75.69 kyr (^{230}Th) (ref. 58). As most samples consist of impure calcite with some degree of detrital contamination, a correction for non-radiogenic ^{230}Th was applied to the measured $^{230}\text{Th}/^{232}\text{Th}$ activity ratio of each sample using an assumed bulk-Earth $^{230}\text{Th}/^{232}\text{Th}$ activity ratio of 0.825 (with a relative error of $\pm 50\%$ and assuming secular equilibrium in the ^{238}U - ^{234}U - ^{230}Th decay chain), as is typical of most other studies⁵⁹. This non-radiogenic ^{230}Th correction reduces the calculated ages of the samples and increases the age uncertainties by an amount dependent on the extent of detrital contamination. All ten dated samples have measured $^{230}\text{Th}/^{232}\text{Th}$ activity ratios of >20 (Supplementary Information section 4), so the detritally-corrected $^{234}\text{U}/^{230}\text{Th}$ ages are only slightly younger than the uncorrected (measured) ages.

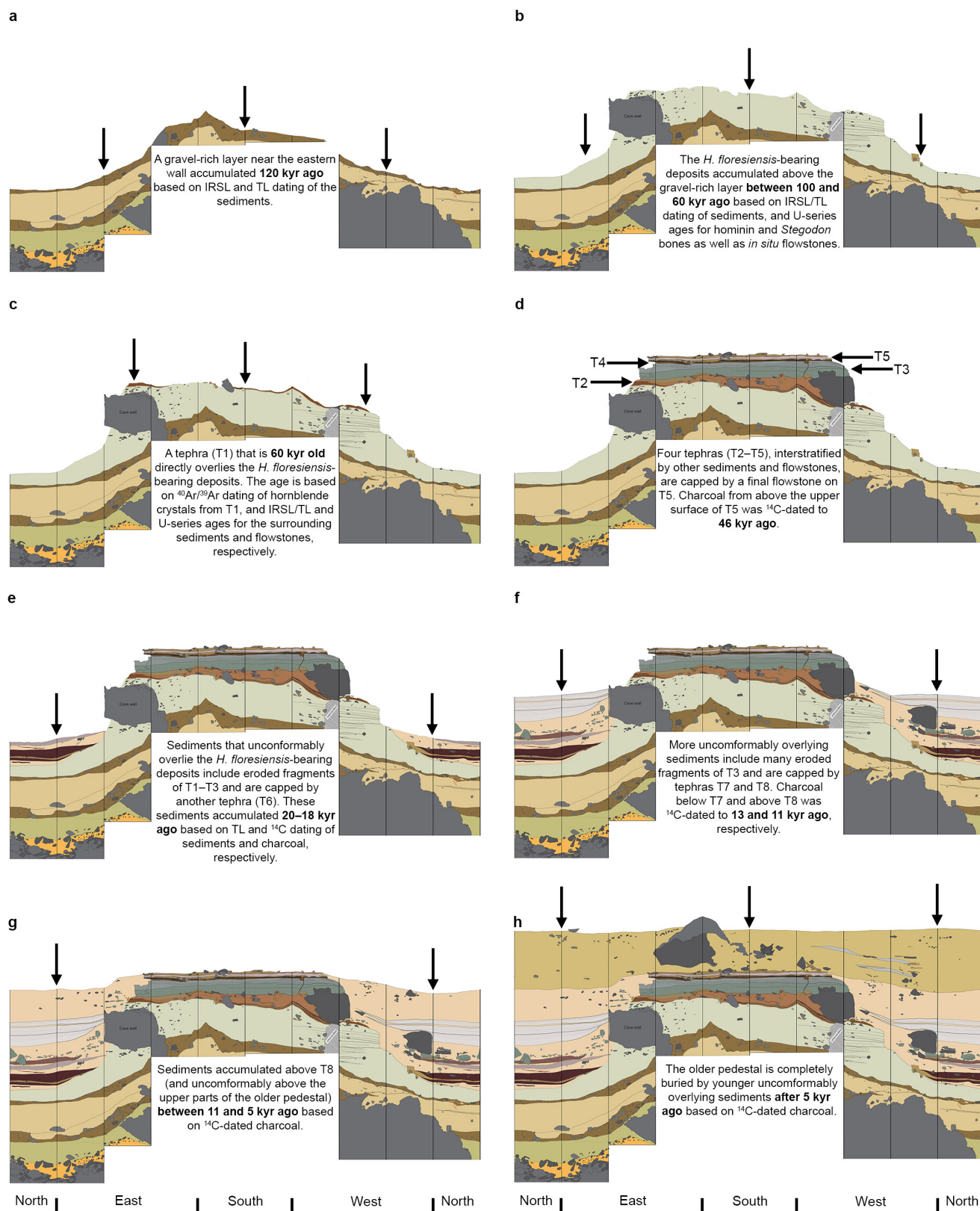
Argon-argon ($^{40}\text{Ar}/^{39}\text{Ar}$) dating of hornblende crystals. Crystals of hornblende were obtained for the 100–150 and 150–250 μm size fractions of the T1 tephra using standard heavy liquid and magnetic separation techniques. Crystals were loaded into wells in 18 mm-diameter aluminium sample discs for neutron irradiation, along with the 1.185 million year-old Alder Creek sanidine²⁶ as the neutron fluence monitor. Neutron irradiation was carried out for 4 min in the cadmium-shielded CLICIT facility at the Oregon State University TRIGA reactor. Argon isotopic analyses of gas released from 6 hornblende aliquots during CO_2 laser step-heating, including a final fusion step ('fuse' in Supplementary Information section 5), were made on a fully automated, high-resolution, Nu Instruments Noblesse multi-collector noble-gas mass spectrometer, using procedures documented previously²⁵. One set of four experiments (Lab IDs 2598 and 2599 in Supplementary Information section 5) consisted of strong degassing of 10 mg hornblende aliquots in square pits in the laser disc, using a beam integrator lens that gives a 'top hat' 6×6 mm energy profile at the focal plane. The laser was then operated at 32 W and this high-temperature, pre-fusion step measured ('D' in Supplementary Information section 5). The fusion step was performed using a conventional focus lens. In the second set of experiments, 20 mg hornblende aliquots were loaded into 5 mm-wide channels in the laser disc and a defocused laser beam was programmed to raster in 50 traverses, each 30 mm in length and separated by 0.1 mm. Low power (<1 W) steps were performed initially to remove loosely trapped argon. At 1 W of laser power (step 'C'), the hornblende crystals began to glow and release significant amounts of ^{39}Ar . This and subsequent steps, including the final fusion, are reported in Supplementary Information section 5 as Lab IDs 2584-03C to -03O (fuse) and 2584-04C to -04I (fuse). Sample gas clean-up was through an all-metal extraction line, equipped with a -130°C cold trap (to remove H_2O) and two water-cooled SAES GP-50 getter pumps (to absorb reactive gases). Argon isotopic analyses of unknowns, blanks and monitor minerals were carried out in identical fashion during a fixed period of 400 s in 14 data acquisition cycles. ^{40}Ar and ^{39}Ar were measured on the high-mass ion counter, ^{38}Ar and ^{37}Ar on the axial ion counter and ^{36}Ar on the low-mass ion counter, with baselines measured every third cycle. Measurement of the ^{40}Ar , ^{38}Ar and ^{36}Ar ion beams was carried out simultaneously, followed by sequential measurement of ^{39}Ar and ^{37}Ar . Beam

switching was achieved by varying the field of the mass spectrometer magnet and with minor adjustment of the quad lenses. Data acquisition and reduction was performed using the program 'Mass Spec' (A. Deino, Berkeley Geochronology Center). Detector intercalibration and mass fractionation corrections were made using the weighted mean of a time series of measured atmospheric argon aliquots delivered from a calibrated air pipette²⁵. The accuracy of the primary air pipette measurements was verified by cross-referencing to data produced from a newly charged second air pipette. Decay and other constants, including correction factors for interference isotopes produced by nucleogenic reactions, are as reported in ref. 25. An isotope correlation (inverse isochron) plot of the data for all 28 aliquots is shown in Extended Data Fig. 3f. The age determined from the inverse isochron is 85 ± 13 kyr for all 28 aliquots, or 79 ± 12 kyr (errors at 1σ) if the data point on the far right-hand side of the plot is excluded. In both cases, the $^{40}\text{Ar}/^{36}\text{Ar}$ intercept value is statistically indistinguishable from the atmospheric ratio of 298.6 ± 0.3 (ref. 60), indicating the absence of significant excess ^{40}Ar in the hornblende crystals.

Radiocarbon (^{14}C) dating of charcoal. Three charcoal samples recovered during excavation were sent to DirectAMS Radiocarbon Dating Service in Bothell, Washington (Supplementary Information section 6). Samples were pretreated using acid-base-acid (ABA) procedures and the ^{14}C content was measured using accelerator mass spectrometry⁶¹. Conventional ^{14}C ages in radiocarbon years before present (BP) were converted to calendar-year age ranges at the 68% and 95% confidence intervals using the SHCal13 calibration data set⁶² and CALIB 7.1 (<http://calib.qub.ac.uk/calib/>). One of these three samples yielded an 'old' radiocarbon age (>40 kyr BP), so we submitted the remaining charcoal to the Oxford Radiocarbon Accelerator Unit (ORAU) for a harsher cleaning protocol known as ABOx-SC^{63,64}. This pretreatment is known to improve charcoal decontamination and has been shown repeatedly to produce more reliable results for 'old' charcoal. However, the harshness of the ABOx-SC procedure often results in large material loss and sample failure, so a new preparative method (AOx-SC) has been developed and tested at the ORAU (K. Douka, personal communication). The AOx-SC procedure does not include a NaOH step and produces identical results to ABOx-SC, but with much higher yields and reduced sample failure. For the Liang Bua sample, ~ 100 mg and ~ 50 mg of hand-picked charcoal underwent ABOx-SC and AOx-SC pretreatments, respectively. Only the AOx-treated charcoal survived the wet chemistry, yielding sufficient carbon for stepped combustion, first at 630°C and then at $1,000^\circ\text{C}$, with the latter fraction collected for graphitization and measurement by accelerator mass spectrometry.

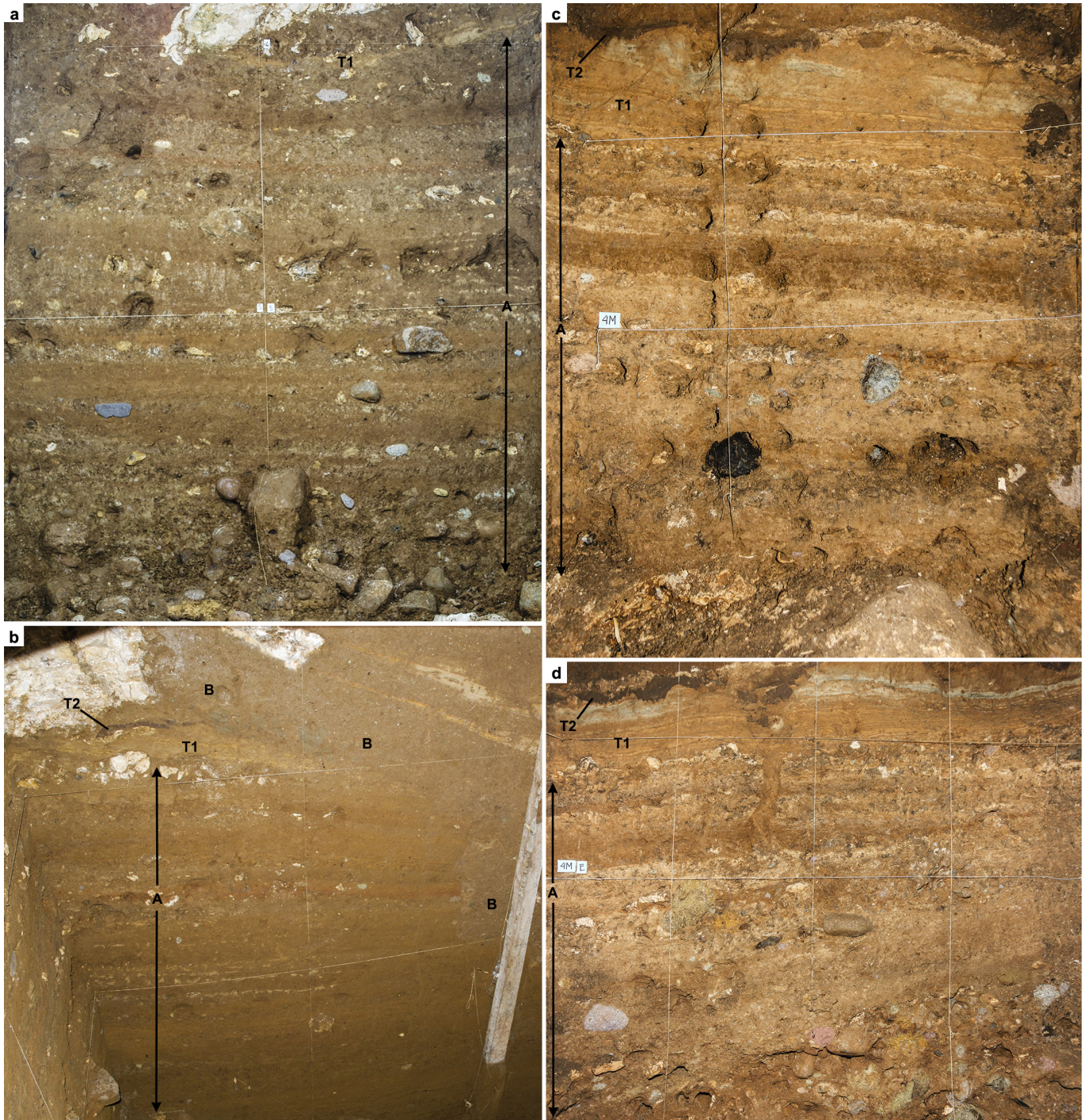
- Heinrich, K. F. J. in *Electron Probe Quantitation* (eds Heinrich, K. F. J. & Newbury, D. E.), 9–18 (Plenum, 1991).
- Jochum, K. P. et al. MPI-DING reference glasses for in situ microanalysis: New reference values for element concentrations and isotope ratios. *Geochim. Geophys. Geosyst.* **7**, Q02008 (2006).
- Alloway, B. V. et al. Correspondence between glass-FT and ^{14}C ages of silicic pyroclastic flow deposits sourced from Maninjau Caldera, west-central Sumatra. *Earth Planet. Sci. Lett.* **227**, 121–133 (2004).
- Jacobs, Z. & Roberts, R. G. An improved single grain OSL chronology for the sedimentary deposits from Diepkloof Rockshelter, Western Cape, South Africa. *J. Archaeol. Sci.* **63**, 175–192 (2015).
- Prescott, J. R. & Hutton, J. T. Cosmic ray contributions to dose rates for luminescence and ESR dating: large depths and long-term time variations. *Radiat. Meas.* **23**, 497–500 (1994).
- Neudorf, C. M., Roberts, R. G. & Jacobs, Z. Sources of overdispersion in a K-rich feldspar sample from north-central India: insights from D_0 , K content and IRSL age distributions for individual grains. *Radiat. Meas.* **47**, 696–702 (2012).
- Neudorf, C. M. *Luminescence Investigations into the Time of Final Deposition of Toba Volcanic Ash and Artefact-bearing Alluvial Sediments in the Middle Son Valley, Madhya Pradesh, India*. PhD thesis, Univ. of Wollongong (2012).
- Huntley, D. J. & Hancock, R. G. V. The Rb contents of the K-feldspars being measured in optical dating. *Anc. TL* **19**, 43–46 (2001).
- Aitken, M. J. *An Introduction to Optical Dating* (Oxford Univ. Press, 1998).
- Bøtter-Jensen, L., Andersen, C. E., Duller, G. A. T. & Murray, A. S. Developments in radiation, stimulation and observation facilities in luminescence measurements. *Radiat. Meas.* **37**, 535–541 (2003).
- Roberts, R. G. et al. Optical dating in archaeology: thirty years in retrospect and grand challenges for the future. *J. Archaeol. Sci.* **56**, 41–60 (2015).
- Li, B. & Li, S.-H. Luminescence dating of K-feldspar from sediments: a protocol without anomalous fading correction. *Quat. Geochronol.* **6**, 468–479 (2011).
- Thiel, C. et al. Luminescence dating of the Stratzing loess profile (Austria) – testing the potential of an elevated temperature post-IR IRSL protocol. *Quat. Int.* **234**, 23–31 (2011).
- Li, B., Roberts, R. G. & Jacobs, Z. On the dose dependency of the bleachable and non-bleachable components of IRSL from K-feldspar: improved procedures for luminescence dating of Quaternary sediments. *Quat. Geochronol.* **17**, 1–13 (2013).
- Huntley, D. J. & Lamothe, M. Ubiquity of anomalous fading in K-feldspars and the measurement and correction for it in optical dating. *Can. J. Earth Sci.* **38**, 1093–1106 (2001).

46. Auclair, M., Lamothe, M. & Huot, S. Measurement of anomalous fading for feldspar IRSL using SAR. *Radiat. Meas.* **37**, 487–492 (2003).
47. Galbraith, R. F., Roberts, R. G., Laslett, G. M., Yoshida, H. & Olley, J. M. Optical dating of single and multiple grains of quartz from Jinmium rock shelter, northern Australia: Part 1, experimental design and statistical models. *Archaeometry* **41**, 339–364 (1999).
48. Galbraith, R. F. & Roberts, R. G. Statistical aspects of equivalent dose and error calculation and display in OSL dating: an overview and some recommendations. *Quat. Geochronol.* **11**, 1–27 (2012).
49. Franklin, A. D., Prescott, J. R. & Robertson, G. B. Comparison of blue and red TL from quartz. *Radiat. Meas.* **32**, 633–639 (2000).
50. Westaway, K. E. The red, white and blue of quartz luminescence: a comparison of D_e values derived for sediments from Australia and Indonesia using TL and OSL emissions. *Radiat. Meas.* **44**, 462–466 (2009).
51. Demeter, F. *et al.* Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc. Natl Acad. Sci. USA* **109**, 14375–14380 (2012).
52. Eggins, S. M. *et al.* *In situ* U-series dating by laser-ablation multi-collector ICPMS: new prospects for Quaternary geochronology. *Quat. Sci. Rev.* **24**, 2523–2538 (2005).
53. Grün, R. *et al.* ESR and U-series analyses of teeth from the palaeoanthropological site of Hexian, Anhui Province, China. *J. Hum. Evol.* **34**, 555–564 (1998).
54. Hellstrom, J. & Pickering, R. Recent advances and future prospects of the U–Th and U–Pb chronometers applicable to archaeology. *J. Archaeol. Sci.* **56**, 32–40 (2015).
55. Zhou, H.-y., Zhao, J.-x., Wang, Q., Feng, Y.-x. & Tang, J. Speleothem-derived Asian summer monsoon variations in Central China, 54–46 ka. *J. Quat. Sci.* **26**, 781–790 (2011).
56. Clark, T. R. *et al.* Discerning the timing and cause of historical mortality events in modern *Porites* from the Great Barrier Reef. *Geochim. Cosmochim. Acta* **138**, 57–80 (2014).
57. Ludwig, K. R. *User's Manual for Isoplot 3.75: a Geochronological Toolkit for Microsoft Excel* (Berkeley Geochron. Center, 2012).
58. Cheng, H. *et al.* The half-lives of uranium-234 and thorium-230. *Chem. Geol.* **169**, 17–33 (2000).
59. Zhao, J.-x., Yu, K.-f. & Feng, Y.-x. High-precision ^{238}U – ^{234}U – ^{230}Th disequilibrium dating of the recent past: a review. *Quat. Geochronol.* **4**, 423–433 (2009).
60. Lee, J.-Y. *et al.* A redetermination of the isotopic abundances of atmospheric Ar. *Geochim. Cosmochim. Acta* **70**, 4507–4512 (2006).
61. Wood, R. From revolution to convention: the past, present and future of radiocarbon dating. *J. Archaeol. Sci.* **56**, 61–72 (2015).
62. Hogg, A. G. *et al.* SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon* **55**, 1889–1903 (2013).
63. Bird, M. I. *et al.* Radiocarbon dating of “old” charcoal using a wet oxidation, stepped-combustion procedure. *Radiocarbon* **41**, 127–140 (1999).
64. Brock, F., Higham, T. F. G., Ditchfield, P. & Bronk Ramsey, C. Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon* **52**, 103–112 (2010).



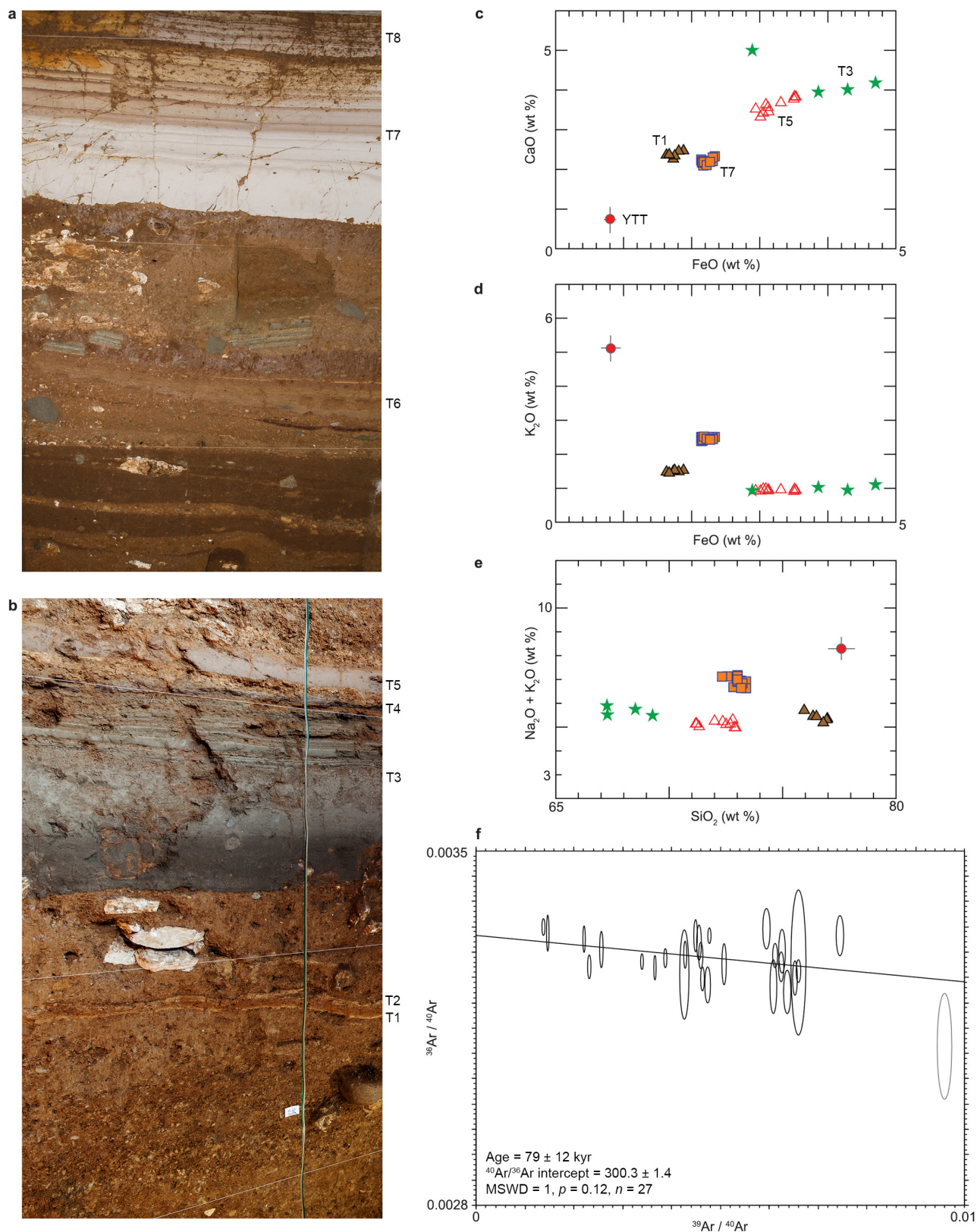
Extended Data Figure 1 | Stratigraphy of the excavated area near the eastern cave wall at eight stages of depositional history, with approximate ages indicated. a–h. Each panel shows the remnant deposits exposed in the 2-m-wide baulks of the following Sectors (from left to right): north VII, east VII, XI and XXIII, south XXIII and XXI, west XXI, XV and XVI, and north XVI. The pedestal deposits shown in **b–d** were

truncated by one or more phases of erosion that resulted in an erosional surface (that is, an unconformity) that slopes steeply down towards the cave mouth (see also Supplementary Video 1). The black arrows relate to the accompanying text in each panel. The maximum depth excavated was 10.75 m in Sector VII (for example, the left two panels in **h**).



Extended Data Figure 2 | Deposits containing the remains of *Homo floresiensis*. These deposits (A) consist of multiple layers of fine-grained sediment interspersed with layers of weathered limestone and loose gravel, and are directly overlain by two tephra (T1 and T2). **a**, South baulk of

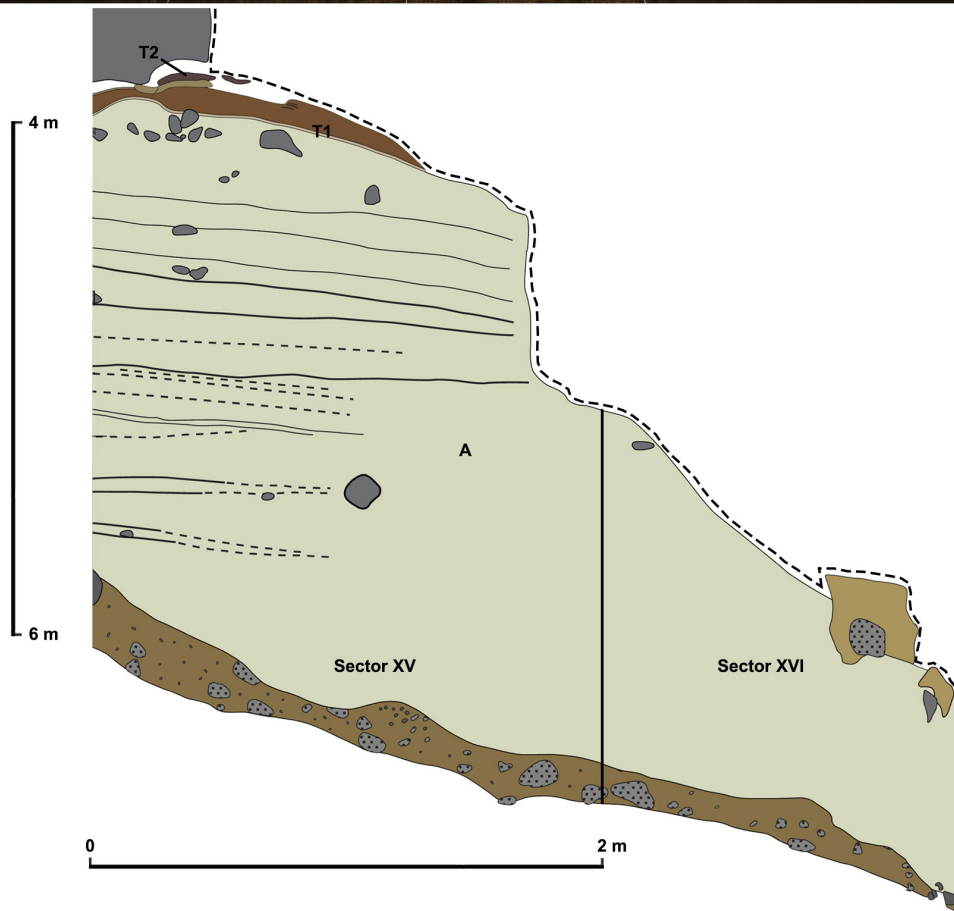
Sector XV, near the eastern cave wall. **b**, West baulk of Sector XV, also showing the unconformably overlying deposits (B). **c**, **d**, North and east baulks of Sector XIX, near the cave centre.



Extended Data Figure 3 | The volcanoclastic deposits at Liang Bua.

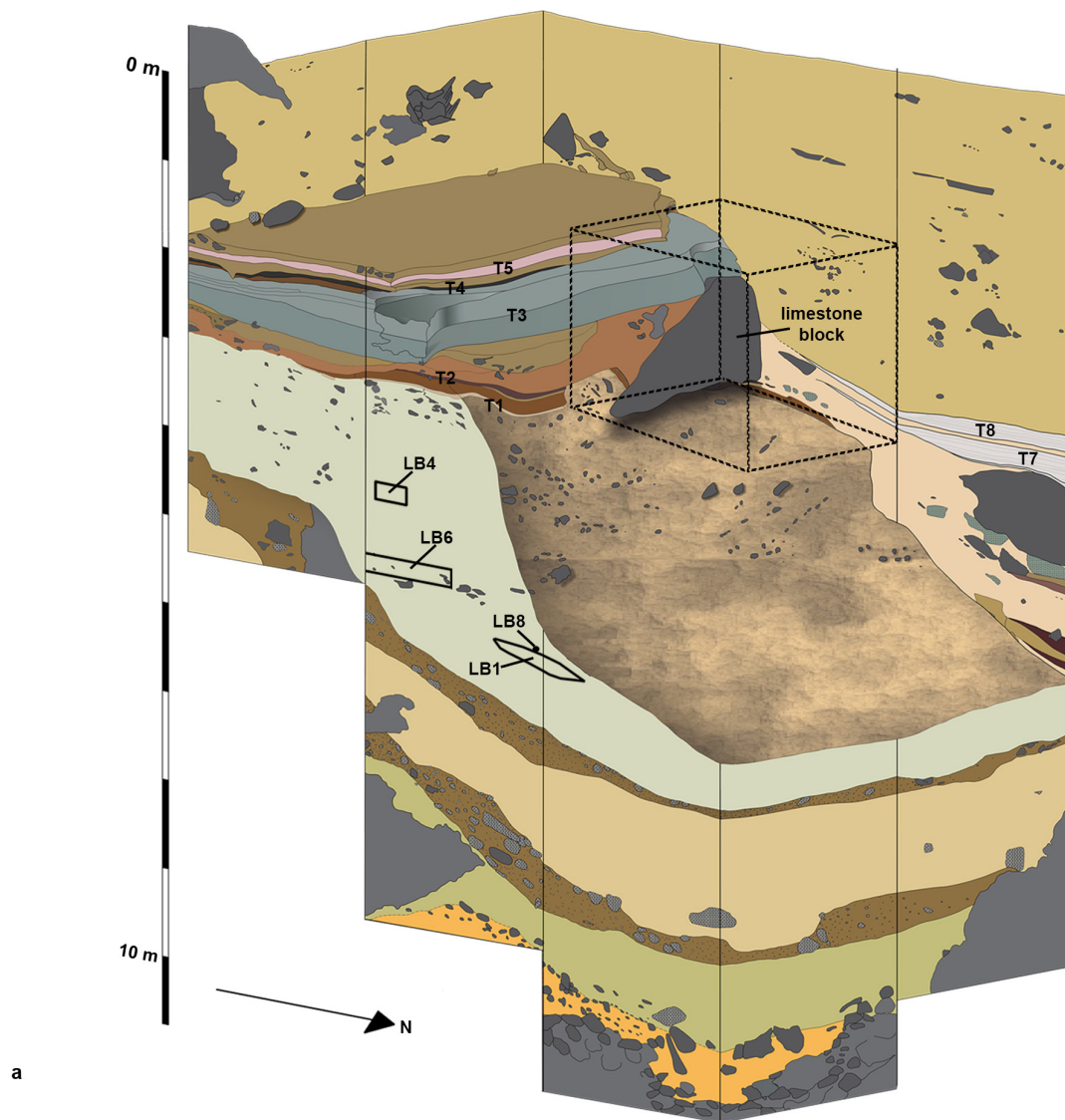
a, Photograph of tephra layers T6–T8 (north baulk of Sector XVI). **b**, Photograph of tephra layers T1–T5 (south baulk of Sector XXI). **c**, Bivariate plot of FeO and CaO concentrations (expressed as weight %), acquired by electron microprobe analysis of glass shards from T1 ($n=6$), T3 ($n=4$), T5 ($n=10$) and T7 ($n=15$), as well as the Youngest Toba Tuff (YTT, $n=207, \pm 1\sigma$)

from northern Sumatra. **d**, Bivariate plot of FeO and K₂O concentrations (symbols as in c). **e**, Bivariate plot of SiO₂ and Na₂O + K₂O concentrations (symbols as in c). **f**, Isotope correlation (inverse isochron) plot for hornblende crystals from T1. The error ellipses represent individual analyses ($n=28$). The ellipse on the far right-hand side was omitted from the $^{40}\text{Ar}/^{39}\text{Ar}$ age determination of 79 ± 12 kyr (at 1σ).



Extended Data Figure 4 | Erosional surface of the pedestal in the west baulks of Sectors XV and XVI. The dashed line marks the steeply sloping boundary between remnant deposits (T2, T1 and the underlying *Homo floresiensis*-bearing sediments) that comprise part of the pedestal (A)

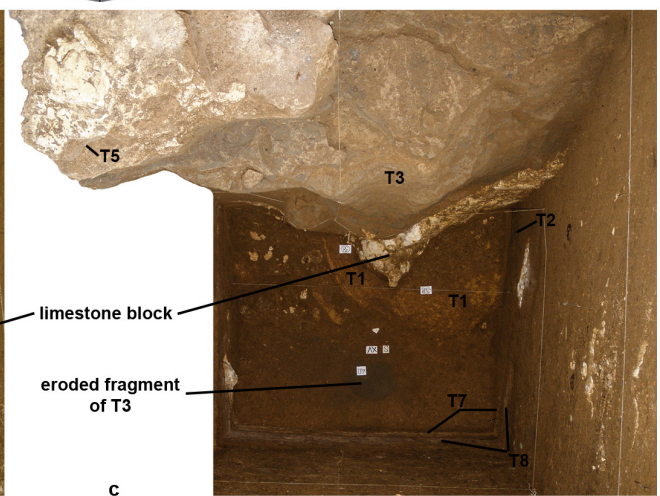
and the much younger deposits (B) that unconformably overlie the contact. **a**, Photograph taken at an upward angle showing the sedimentary differences between the deposits above and below the erosional boundary. **b**, Illustration of the erosional surface and underlying deposits shown in **a**.



a



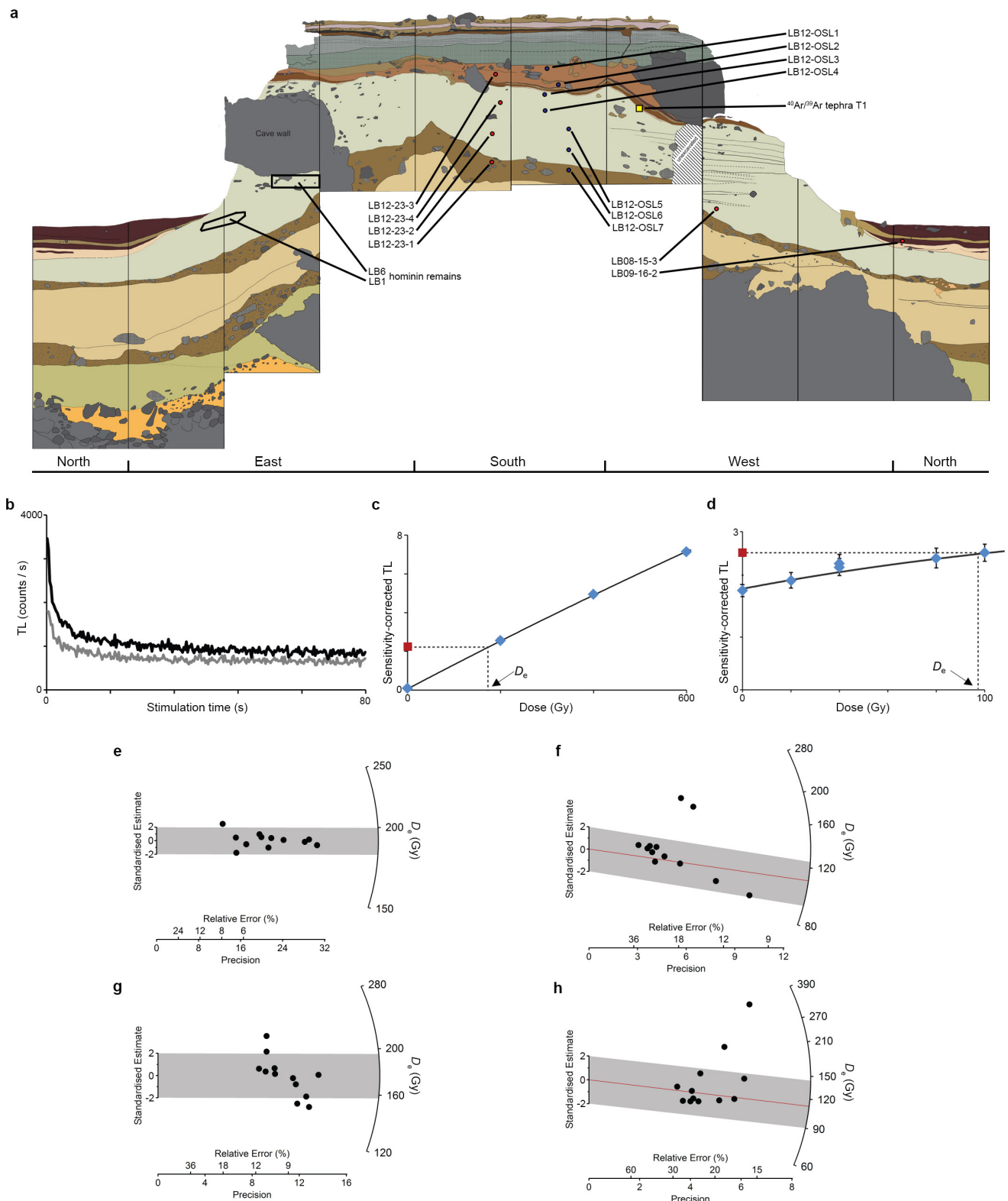
b



c

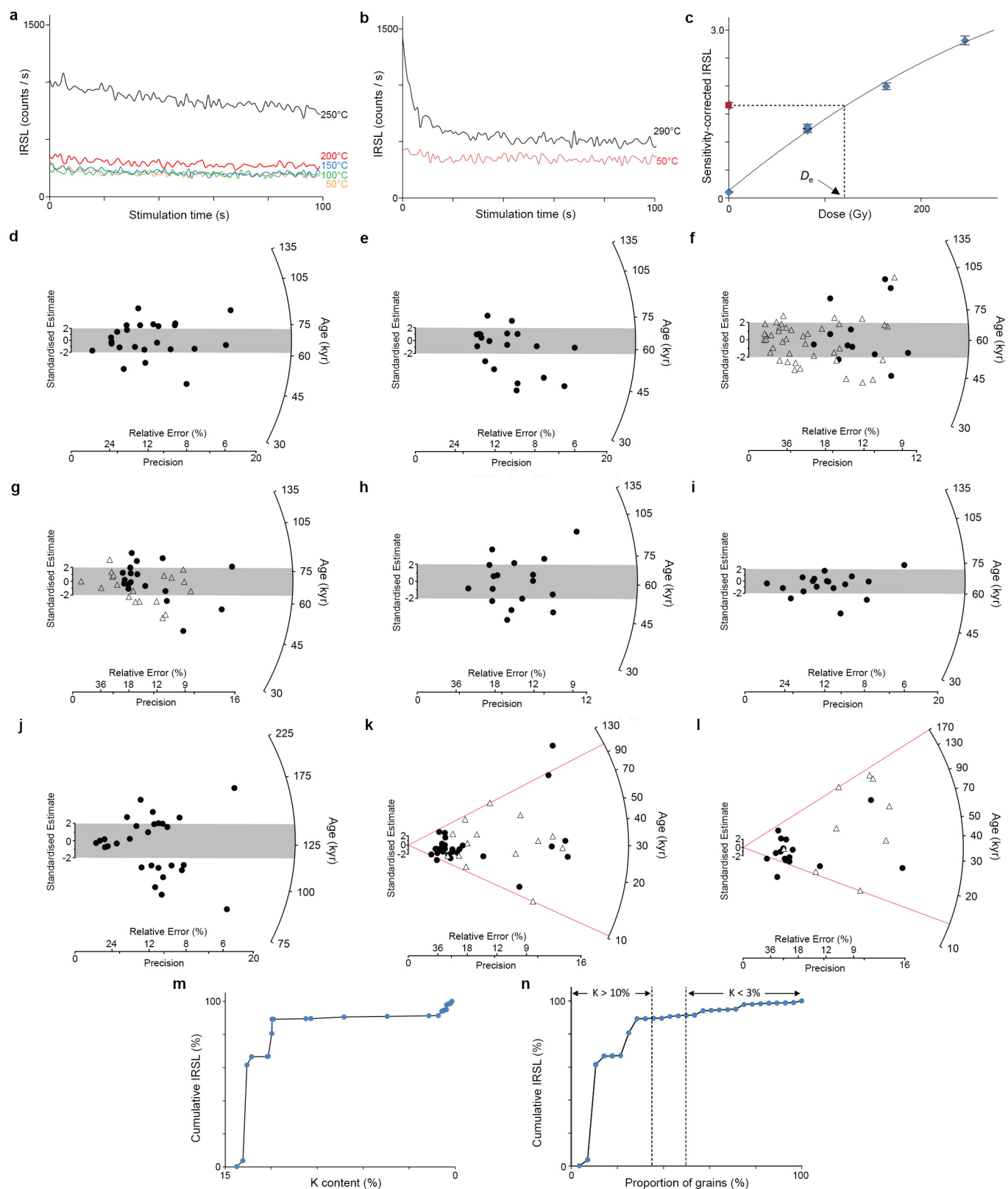
Extended Data Figure 5 | Erosional surface of the pedestal near the eastern wall of the cave. a, Illustration of the erosional surface and the locations of LB1, LB4, LB6 and LB8 below the boundary (see also Fig. 3). The deposits that unconformably overlie the pedestal are shown in the

south and west baulks. The stippled cube outlines the photographed area (in Sector XV) shown in b and c. Both photographs taken from above, with north towards the bottom of the page.



Extended Data Figure 6 | Locations of sediment samples dated in this study and TL data for quartz grains from Liang Bua. a, Stratigraphy of the excavated area near the eastern cave wall (Sector baulks as in Extended Data Fig. 1) with TL samples indicated by red circles, IRSL samples by blue circles and the $^{40}\text{Ar}/^{39}\text{Ar}$ sample by a yellow square. Also shown are the TL and IRSL sample codes and the locations of hominin remains LB1 and LB6. **b**, Representative isothermal (260 °C) TL decay curves for the natural (black line) and test dose (grey line) signals from sample LB08-15-3. **c**, **d**, Regenerated TL dose-response curves for one pair of Aliquots A and B of sample LB08-15-3, respectively; the equivalent dose (D_e) is estimated by

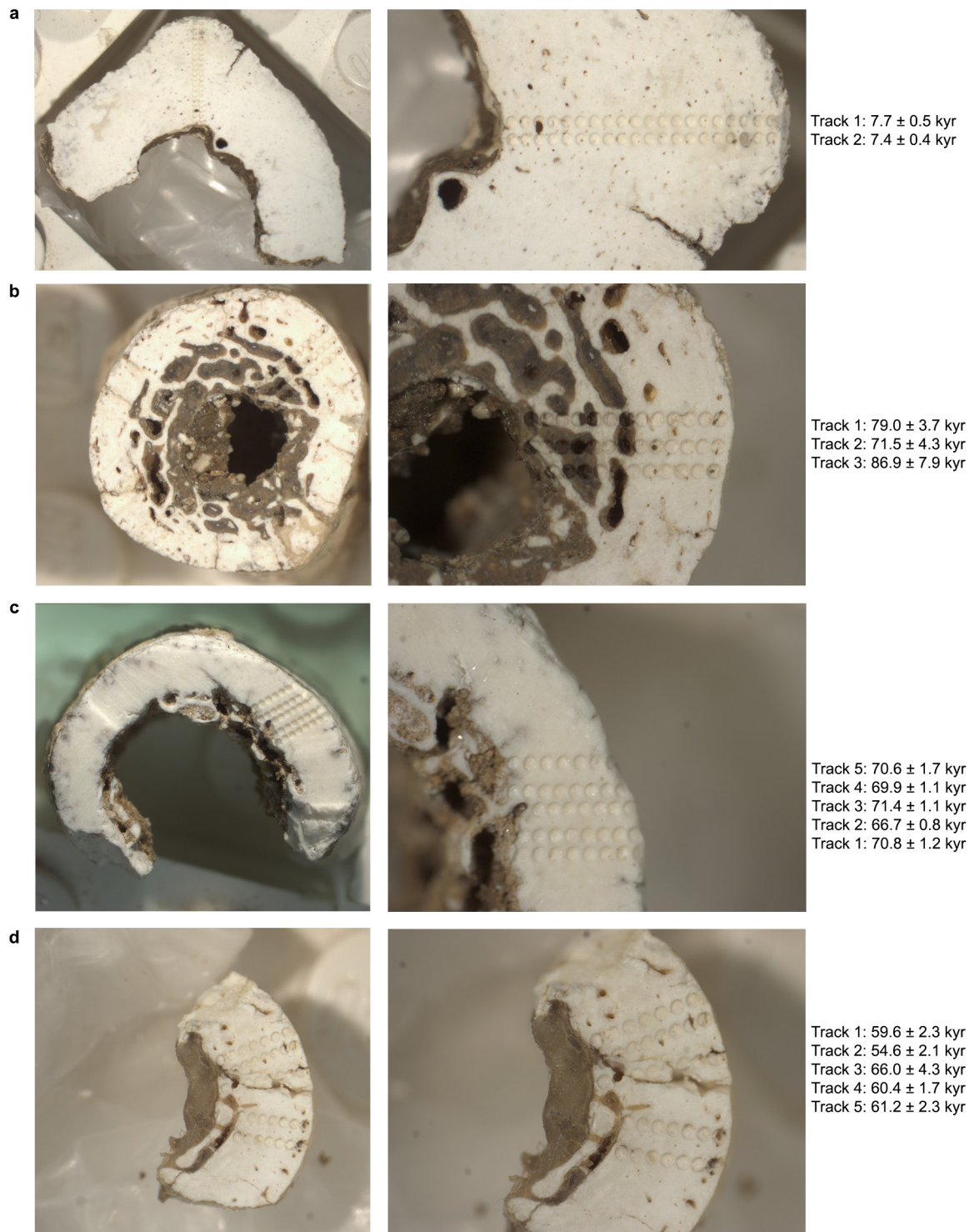
projecting the natural signal (red square) on to the dose-response curve fitted to the regenerated signals (blue diamonds). **e**, Radial plot^{47,48} of D_e values for Aliquot A ($n = 12$) of sample LB08-15-3; the grey band is centred on the weighted mean D_e calculated using the central age model. **f**, Radial plot of the corresponding D_e values for Aliquot B ($n = 12$) of the same sample. The grey band is centred on the central age model estimate, with the two high- D_e outliers omitted. The red line intersects the right-hand axis at the D_e calculated by fitting the minimum age model^{47,48} to all 12 values. **g**, **h**, Radial plots of D_e values for Aliquots A and B of sample LB12-23-1 (symbols as in **e** and **f**).



Extended Data Figure 7 | See next page for caption.

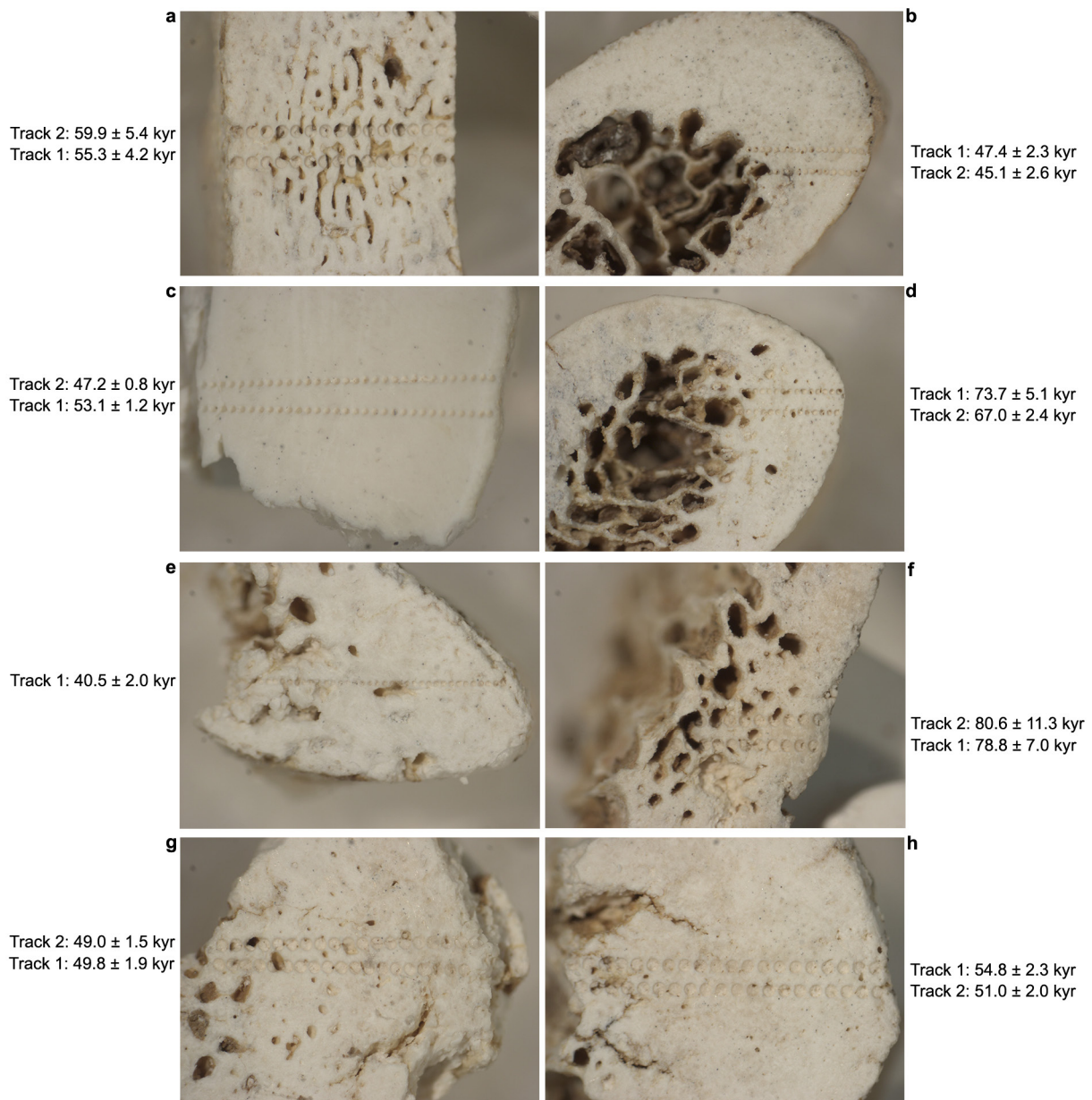
Extended Data Figure 7 | IRSL data and potassium (K) concentrations for feldspar grains from Liang Bua. **a**, Representative IRSL (50 °C) and multiple elevated temperature (100–250 °C) post-infrared IRSL (pIRIR) decay curves for a single aliquot of sample LB12-OSL1. **b**, IRSL (50 °C) and pIRIR (290 °C) decay curves for a different aliquot of LB12-OSL1. **c**, Regenerated pIRIR (290 °C) dose–response curve for the aliquot shown in **b**; the equivalent dose (D_e) is estimated by projecting the natural signal (red square) on to the dose–response curve fitted to the regenerated signals (blue diamonds). **d–j**, Radial plots of IRSL ages (corrected for residual dose and anomalous fading) for single aliquots of each sample: **d**, LB12-OSL1; **e**, LB12-OSL2; **f**, LB12-OSL3; **g**, LB12-OSL4; **h**, LB12-OSL5; **i**, LB12-OSL6; and **j**, LB12-OSL7. IRSL ages were also obtained for single grains of samples LB12-OSL3 and LB12-OSL4, and are shown as open triangles in **f** and **g**. The grey bands in each plot are centred on the

weighted mean ages calculated using the central age model. **k**, **l**, Radial plots of IRSL ages (corrected as for **d–j**) for samples LBS7-40a and LBS7-42a, respectively; single aliquots are shown as filled circles and single grains as open triangles. The upper and lower red lines intersect the right-hand axis at the maximum and minimum single-grain ages, respectively. **m**, Distribution of pIRIR intensities from 28 individual grains of feldspar from sample LB12-OSL3 that had been given a regenerative dose of 80 Gy. The relative contribution of each grain to the total (cumulative) pIRIR light sum is plotted as a function of K concentration (measured by wavelength-dispersive X-ray spectroscopy); note the reversed scale on the x -axis. **n**, Cumulative pIRIR light sum for the same 28 grains as shown in **m**, plotted as a function of grains ranked by K concentration (which decreases from left to right).



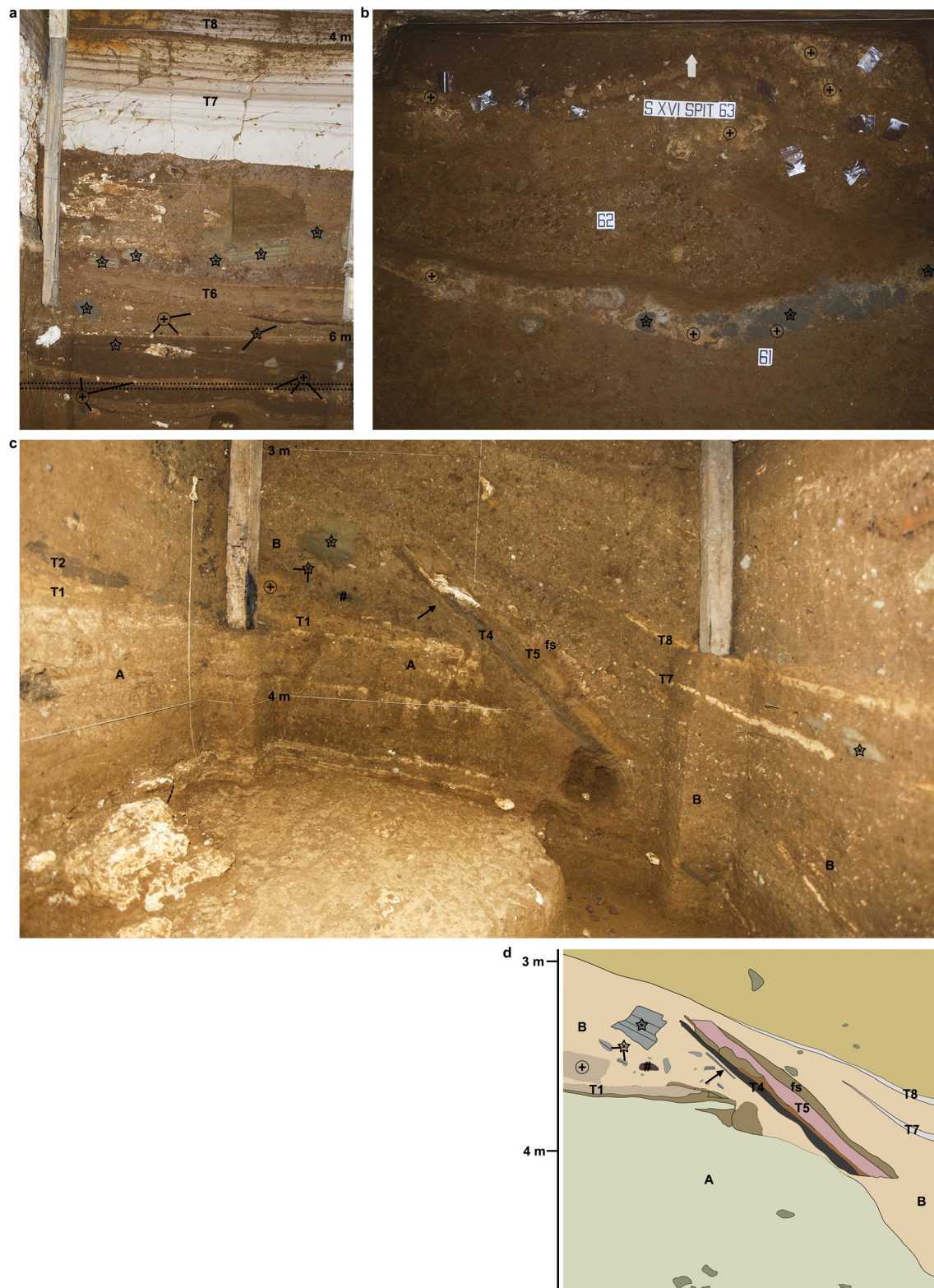
Extended Data Figure 8 | Laser-ablation uranium-series analyses of hominin bone fragments from various Sectors and spits (depth intervals), and their modelled ages. a, Modern human femur (132A/LB/27D/03) from Sector IV, spit 27 (265–275 cm). b, *Homo floresiensis*

ulna (LB1/52) from Sector XI, spit 58A (575–585 cm). c, *H. floresiensis* ulna (LB2/1) from Sector IV, spit 42D (415–425 cm). d, *H. floresiensis* ulna (LB6/3) from Sector XI, spit 51 (505–515 cm). Each laser spot is $265\mu\text{m}$ in diameter and the age errors are at 2σ .



Extended Data Figure 9 | Laser-ablation uranium-series analyses of bone fragments of *Stegodon florensis insularis* from various spits (depth intervals) in Sector XI, and their modelled ages. a, U-s-01/LB/XI/32/04, spit 32 (315–325 cm). b, U-s-02/LB/XI/45/04, spit 45 (445–455 cm). c, U-s-03/LB/XI/47/04, spit 47 (465–475 cm). d, U-s-04/LB/XI/49/04,

spit 49 (485–495 cm). e, U-s-05/LB/XI/51/04, spit 51 (505–515 cm). f, U-s-06/LB/XI/52/04, spit 52 (515–525 cm). g, U-s-07/LB/XI/65/04, spit 65 (645–655 cm). h, U-s-08/LB/XI/65B/04, spit 65B (645–655 cm). Each laser spot is $265 \mu\text{m}$ in diameter and the age errors are at 2σ .



Extended Data Figure 10 | Deposits stratigraphically above the unconformity in Sector XVI and displaced slab of deposit in Sector XXII. **a**, The north baulk (~2 m wide) of Sector XVI. **b**, Excavated floors (white arrow points north) of spits 61–63 (615–635 cm depth); the field of view is ~1.6 m in width. The stippled box in **a** indicates the floor of spit 63 in **b**, where fragments of T1 (+) are visible in spit 63, and fragments of T3 (*) and T1 are concentrated in the band just above the label for spit 61. Eroded fragments (between about 1 cm and 60 cm in size) of T1, T2 and T3 have been consistently recovered from deposits unconformably overlying the erosional surface of the pedestal, indicating reworking of the pedestal deposits before ~13 kyr cal. BP. **c**, Photograph of the west baulk

and parts of the south and north baulks (at left and right, respectively) of Sector XXII showing a displaced slab of deposit that contains intact portions of the uppermost part of T3 (arrow) and the overlying layers, up to and including the flowstone (fs) that caps T5. The stratigraphic position of the slab beneath T7 and T8 indicates that it broke away from its original location, slightly to the south, and slid down the steeply sloping erosional surface before ~13 kyr cal. BP. Also shown are the *Homo floresiensis*-bearing deposits (A) and the unconformably overlying deposits (B), which include eroded fragments of T1 (+), T2 (#) and T3 (*). **d**, Illustration of the west baulk of Sector XXII, as shown in **c**.

Anatomy and function of an excitatory network in the visual cortex

Wei-Chung Allen Lee¹, Vincent Bonin^{1,2}, Michael Reed¹, Brett J. Graham¹, Greg Hood³, Katie Glattfelder⁴ & R. Clay Reid^{1,4}

Circuits in the cerebral cortex consist of thousands of neurons connected by millions of synapses. A precise understanding of these local networks requires relating circuit activity with the underlying network structure. For pyramidal cells in superficial mouse visual cortex (V1), a consensus is emerging that neurons with similar visual response properties excite each other^{1–5}, but the anatomical basis of this recurrent synaptic network is unknown. Here we combined physiological imaging and large-scale electron microscopy to study an excitatory network in V1. We found that layer 2/3 neurons organized into subnetworks defined by anatomical connectivity, with more connections within than between groups. More specifically, we found that pyramidal neurons with similar orientation selectivity preferentially formed synapses with each other, despite the fact that axons and dendrites of all orientation selectivities pass near ($<5\mu\text{m}$) each other with roughly equal probability. Therefore, we predict that mechanisms of functionally specific connectivity take place at the length scale of spines. Neurons with similar orientation tuning formed larger synapses, potentially enhancing the net effect of synaptic specificity. With the ability to study thousands of connections in a single circuit, functional connectomics is proving a powerful method to uncover the organizational logic of cortical networks.

Pyramidal cells in the rodent primary visual cortex (V1) respond to highly specific visual features, resulting in diverse receptive field preferences, in contrast with the less selective responses of most inhibitory neurons^{6–13}. A model is emerging in which V1 responses arise from the selective amplification of thalamocortical signals^{1,3,4} through recurrent inputs from other pyramidal neurons^{2,5,14,15}. Evidence for functionally specific cortical amplification has been seen in physiological^{2,3,14} and optogenetic⁴ studies; however, firm anatomical evidence at the synaptic level has been lacking (but see ref. 5). A recent study also showed that neurons with similar orientation preference share stronger connections¹⁴; however, it is unknown if this effect is due to stronger synapses, more synapses, or perhaps spatially clustered synapses.

To test these hypotheses, we measured the receptive-field properties and reconstructed the detailed anatomy of the same visual cortical neurons, identifying actual synapses versus axonal-dendritic appositions⁸ (Fig. 1). We combined *in vivo* cellular resolution optical imaging with *ex vivo* electron microscopy (EM) reconstructions. We measured cellular calcium responses, which reflect the firing of action potentials, using the genetically encoded indicator GCaMP3 to characterize the sensory responses of $\sim 300\mu\text{m} \times 300\mu\text{m} \times 200\mu\text{m}$ volume of an awake, behaving mouse¹⁰ (Fig. 1b–c). Visual stimuli consisted of drifting sinusoidal gratings of different spatial and temporal frequencies, orientations, and directions (Fig. 1b and Extended Data Fig. 1). In addition to cell bodies in layers 2/3 (L2/3), we measured signals (Extended Data Fig. 2) from large calibre apical dendrites that continued beyond the depth of our imaging volume and had branching morphologies consistent with deep layer pyramidal cells (see Methods). From their

responses, we estimated the peak preferred orientation for each cell (Extended Data Fig. 3), with neurons' visual preferences typically maintained across 12 days of chronic imaging (Extended Data Fig. 1).

After locating the functionally imaged region using vascular landmarks (Extended Data Fig. 4 and Supplementary Video 1), we cut a series of $\sim 3,700$ serial EM sections, which were imaged with a transmission electron microscope camera array⁸ (TEMCA) at $\sim 4\text{ nm} \times 4\text{ nm} \times 40\text{ nm}$ per voxel. The EM-imaged region spanned $450\mu\text{m} \times 450\mu\text{m} \times 150\mu\text{m}$, consisting of ~ 10 million camera images and ~ 100 TB of raw data. We traced the processes of excitatory pyramidal neurons located within the middle third of the EM volume ($450\mu\text{m} \times 450\mu\text{m} \times 50\mu\text{m}$, Supplementary Videos 2 and 3) using software (CATMAID¹⁶) allowing distributed annotation of large image data sets. Teams of trained annotators traced and validated wire-frame models of the dendritic and axonal arbors of neurons selected for reconstruction (chosen because they exhibited visual responses), and located all outgoing synapses along the annotated axons (Fig. 1d, and Extended Data Figs 5 and 6a). For each synapse, we traced the postsynaptic dendrite centripetally until they reached either the cell body or the boundary of the aligned EM volume (Extended Data Fig. 6b, and Supplementary Data 1–3). We did not retrogradely trace axons providing input to selected neurons because of the lower probability that they originated from cells in the volume. Henceforth, we limit our discussion to the excitatory network from selected pyramidal cells onto spines of excitatory targets.

We first examined the structure of the network broadly, by examining the connectivity of 1,278 reconstructed neuronal targets (Table 1). The core of the network (201 neurons connected with more than one other; that is, degree ≥ 2 , no leaf nodes) exhibited a modular structure (Fig. 1e), quantified with a modularity index (Q , which can range from -0.5 to 1) that measures the degree to which a network can be divided into groups, with positive values indicating higher levels of connectivity within than between groups¹⁷. Modularity in the reconstructed network was significantly higher than would be expected by chance ($Q_{\text{mean}} = 0.55 \pm 0.003$ vs 0.50 ± 0.009 , mean \pm s.d., $P \approx 0$, permutation test; Fig. 1e and Extended Data Fig. 7a–d, see Methods) and was not a consequence of higher order motifs (Extended Data Fig. 7). Modularity of the network offers positive evidence for pyramidal cell subnetworks that have previously been inferred from physiological recordings^{2,18,19}.

We next examined the degree to which network connectivity reflects the neurons' sensory properties (Figs 1f and 2). We focused on 46 L2/3 excitatory neurons and 4 deep layer apical dendrites that showed reproducible responses to visual stimulation (trial correlations for spatial and temporal frequency or stimulus position experiments, $P_{\text{stfr}} < 0.05$ and $P_{\text{pos}} < 0.01$, respectively, excluding cells with stimulus edge effects, see Methods) and identified their postsynaptic targets. Within the reconstructed network, 43 functionally characterized neurons made 990 synapses, 443 of which were onto inhibitory interneurons and 547 onto other pyramidal cells (Table 1). The large fraction of synapses onto

¹Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Neuro-Electronics Research Flanders, a research initiative by imec, Vlaams Instituut voor Biotechnologie (VIB) and Katholieke Universiteit (KU) Leuven, 3001 Leuven, Belgium. ³Biomedical Applications Group, Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. ⁴Allen Institute for Brain Science, Seattle, Washington 98103, USA.

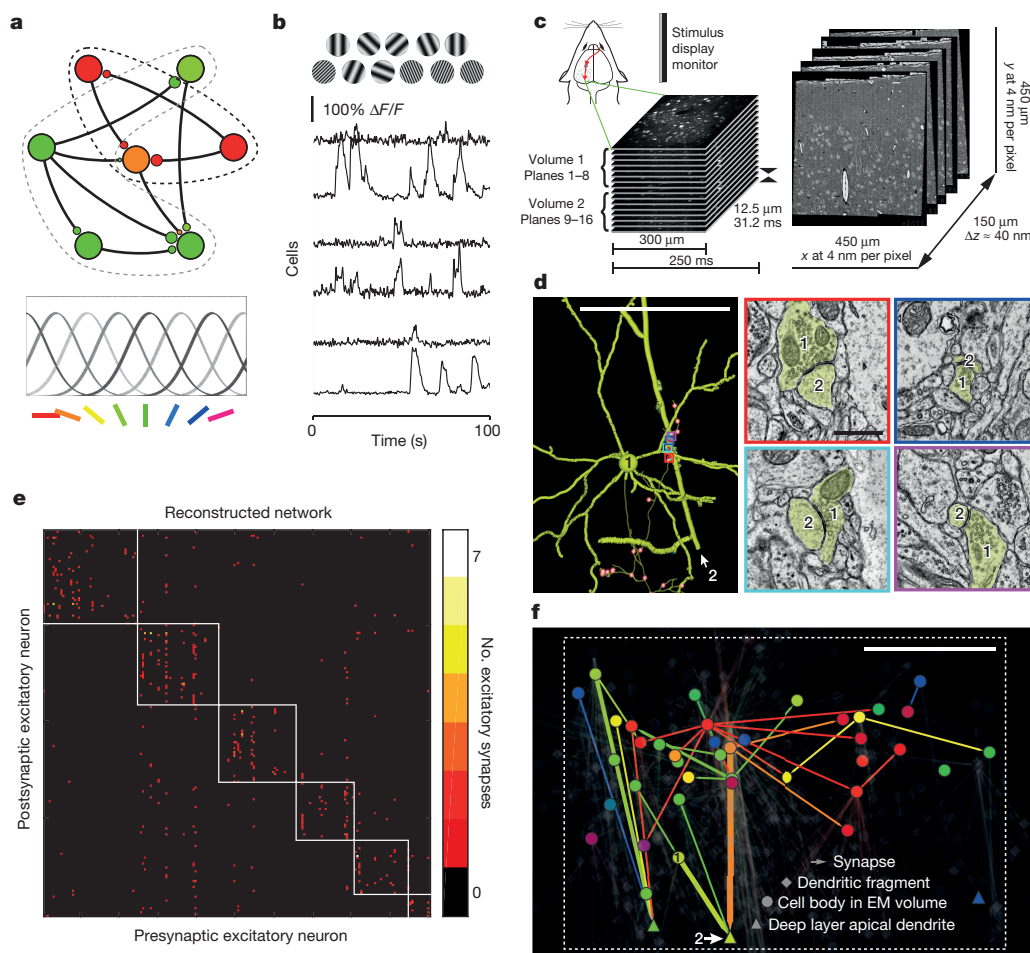


Figure 1 | Functional organization of cortical excitatory network connectivity. **a**, Schematic representation of functionally selective connections between excitatory neurons. Excitatory pyramidal cells (large circles) with different preferred orientations provide synaptic input (smaller circles) to one another (bottom, colour code used throughout to indicate stimulus orientation that evokes peak physiological responses). **b**, Example stimuli (top) and time courses of $\Delta F/F$ signals from single cells. **c**, Combined *in vivo* two-photon calcium imaging and electron microscopy (EM). Top left, schematic of imaging in the awake mouse, targeting the monocular region of the primary visual cortex (largest dotted-outlined region surrounded by higher visual areas). Red arrows represent the visual pathway from eye to visual thalamus to cortex. Serial EM sections (right) were cut orthogonal to the functional imaging planes (left). **d**, Left, reconstruction of a layer 2/3 (L2/3) pyramidal neuron (cell 1) synapsing onto a deep layer apical dendrite (cell 2), reconstructed from EM (colour denotes preferred orientation of neuron). The thinnest process is the axon, dendrites of the L2/3 neuron are thicker, and the deep layer apical dendrite is rendered with the largest calibre. Dendritic spines were traced only if they participated in connections between reconstructed neurons. Synapses are shown as small spheres. Presynaptic

inhibitory neurons is consistent with past observations^{8,20} and seems to be a characteristic of L2/3 in the mouse visual cortex. Consistent with physiological studies, connectivity between excitatory neurons was more likely for pairs with similar orientation preferences² (Fig. 2b, $P < 0.05$, Cochran–Armitage test). This preferential connectivity is not explained by the anatomical arrangement of the cell bodies whose function showed no discernible dependence on distance (Extended Data Fig. 8).

As axons and dendrites of physiologically characterized neurons were traced exhaustively within the volume, we could further evaluate in an unbiased fashion a potential role of axonal and dendritic geometry in specific connectivity. We used a modified measure of potential synapses²¹, which our reconstructions allowed us to compute directly.

boutons not connecting this cell pair are smaller red spheres with white centres. EM micrographs of synapses (right) with similar overlay colours corresponding to the very similar peak orientation preferences for cells 1 and 2. **e**, Connectivity matrix of 201 excitatory neuronal targets in our network reconstruction with multiple synaptic partners (that is, degree ≥ 2 , no leaf nodes). Colour represents the number of synapses (colour key, right). Subnetworks of interconnected neurons (white boxes) detected using a consensus method of Louvain clustering ($Q = 0.55 \pm 0.003$, mean \pm s.d.)^{17,31}. **f**, Network graph of functionally characterized pyramidal neurons and their connections to other excitatory neuronal targets (transparent) viewed coronally. Arrows represent synaptic connections and their shaft thickness is scaled by number of synapses (range: 1–7). Neuronal targets with cell bodies in the EM volume drawn as circles, large calibre apical dendrites that exit the volume (deep layer pyramidal cells) drawn as triangles, and other postsynaptic targets (dendritic fragments) drawn as diamonds. Nodes are positioned by cell body location, synapse location (for dendritic fragments) or, for deep layer apical dendrites, by the deepest position when exiting the volume. Cells labelled 1 and 2 are the same as in **d**. Bounding box matches region in Extended Data Fig. 6. Scale bars, **d**, left, 100 and right, 1 μm ; **f**, 100 μm .

For each pair of neurons, we quantified potential synapse length, or the length that each dendrite (L_d) travelled within a distance (s) from a given axon, where s can be considered a maximum distance a spine could reach to make a connection ($s = 5 \mu\text{m}$: Fig. 2a, c; $s = 1 \mu\text{m}$: Extended Data Fig. 9; see Methods). This helps us formulate a fine-scale version of Peters' rule: the hypothesis that the probability of a connection is proportional to the degree of axo-dendritic proximity, independent of other factors. As expected, connected neurons had significantly more L_d than unconnected pairs, that is, their geometry provided more opportunities to make synapses ($P \approx 0$, permutation tests, $s = 5 \mu\text{m}$, Fig. 2c; $s = 1 \mu\text{m}$: Extended Data Fig. 9). More interestingly, by considering potential synapse length between pairs with different relative orientations, we found that the neuropil is not functionally

Table 1 | Cortical network reconstruction

Excitatory neuronal targets	1,278
Presynaptic neurons	130
Postsynaptic partners	1,228
Deep layer apicals	91
Connected layer 2/3 neurons to deep layer apical pairs	185
Inhibitory neuronal targets	581
Functionally characterized neurons	50
Characterized connected neurons	43
Characterized deep layer apicals	4
Characterized neurons connected to deep layer apicals	20
Functionally characterized connected pairs	29
Unique characterized presynaptics	15
Unique characterized postsynaptics	21
Characterized pairs containing deep layer apicals	8
Unique characterized deep layer apicals in pairs	2
Synapses from characterized neurons	990
Synapses onto inhibitory targets	443
Synapses onto excitatory targets	547
Reconstructed PSDs between characterized pairs	39
Multi-synapse connected pairs	115
Multi-synapse connected pairs with deep layer apicals	62
Multi-synapse connected presynaptic neurons	51
Multi-synapse connected postsynaptic non-deep layer apicals	52
Multi-synapse connected postsynaptic deep layer apicals	41

Numbers of reconstructed neuronal targets (neurons and dendritic fragments), synaptically connected neuron pairs, and synapses analysed in the EM data set.

organized ($P > 0.5$, permutation tests relative to a uniform distribution, $s = 5 \mu\text{m}$: Fig. 2d, black line, and 2e; $s = 1 \mu\text{m}$: Extended Data Fig. 9). Therefore, the preferential connectivity (Fig. 2d, red versus blue line and red versus black line both, $P < 0.05$, permutation tests; and Fig. 2f, $P < 0.05$, Cochran–Armitage test) between neurons with similar orientation preferences must be the result of mechanisms that take place at the scale of spines: ~ 1 to $5 \mu\text{m}$. Because synapses onto the apical dendrites of deep layer neurons might follow different rules^{5,22}, we also confirmed statistical significance ($P < 0.05$, permutation test, data not shown) with only L2/3 connections.

Recent work suggests that, along with connection probability, the amplitude of excitatory postsynaptic potentials (EPSPs) correlates with receptive field properties¹⁴. Mechanisms underlying the strength of these unitary responses are unknown, for instance whether they are due to multiple synapses, the spatial organization of their synapses, or stronger synapses. Although our sample size was not sufficient to examine the relationship between multiple synapses and function, past work demonstrated that virtually all pyramidal cells are connected by multiple synapses (reviewed in ref. 23). We therefore examined whether synapses connecting pairs of neurons were spatially clustered (Fig. 3a, Fig. 1d, f, thick arrows, and Supplementary Data 3) in a population of 51 reconstructed presynaptic neurons connected by multiple synapses onto one or more postsynaptic targets (Table 1). We first computed the synapse rate (λ), or synapses per $100 \mu\text{m}$ of L_d . As expected, connected neurons coupled by multiple synapses occur more frequently than predicted by chance, having far higher synapse rates for additional synapses compared to a Poisson model ($P \approx 0$, permutation tests relative to a Poisson process with λ , Fig. 3b, compared to Fig. 2f, see Methods). Next, we found that pyramidal neurons were frequently interconnected by multiple synapses arranged close together ($17.2 \pm 2.9 \mu\text{m}$, median \pm s.e.m., Fig. 3c) consistent with recent analyses of axon fragments in L5 apical dendrites²⁴ and hippocampus²⁵. A careful analysis of synapse locations considering potential synapse length (L_d), however, shows that closely spaced synapses are not specifically enriched. By comparing the relative magnitude of L_d for regions

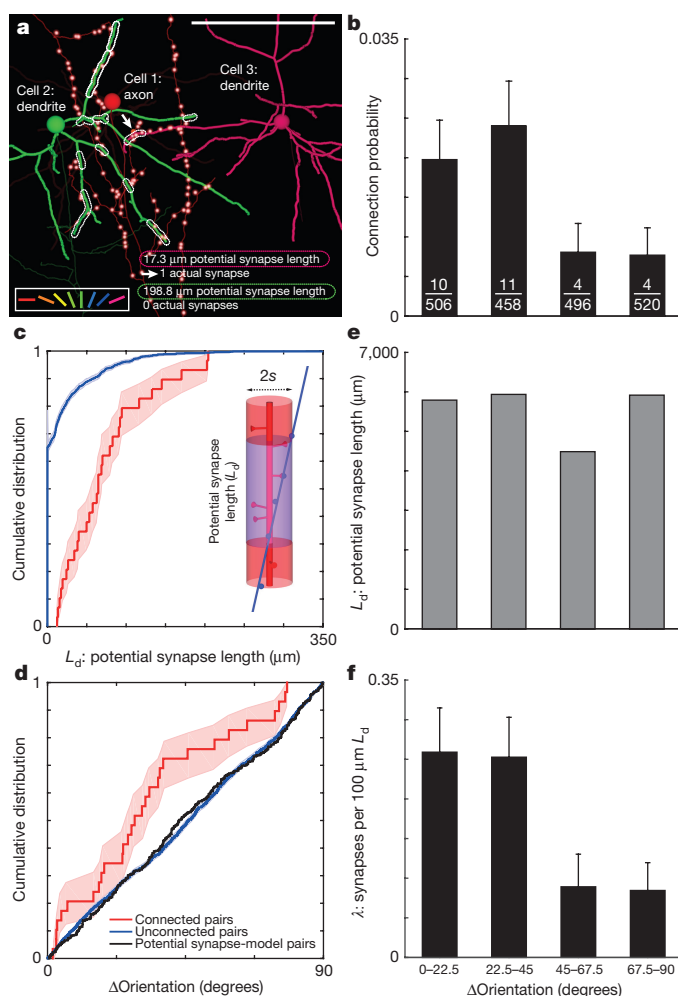


Figure 2 | Synaptic connectivity between pyramidal cells predicted by function, over and above proximity. **a**, Reconstructions of three neurons with regions of potential synaptic connectivity (L_d , dotted outlines), which quantify the dendritic path length of cells 2 and 3 within a maximal spine length ($s = 5 \mu\text{m}$) of the axon of cell 1. Neurons colour-coded by preferred stimulus orientation (colour key, bottom left; same as Fig. 1). Note, the large L_d , but lack of actual synapses between cells 1 and 2 and small L_d between cells 1 and 3 where an actual synapse (arrow) is observed along the L_d . The axon of the postsynaptic cells and dendrites of the presynaptic cell are transparent. Only presynaptic boutons (small red spheres with white centres) from the presynaptic cell are visible. **b**, Synaptically connected neurons have similar orientation preference. Connection probability as function of orientation preference difference across the reconstructed population ($P < 0.05$, Cochran–Armitage test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$). Bins include pairs of neurons with differences in preferred orientation of 0° to 22.5° , 22.5° to 45° , 45° to 67.5° , and 67.5° to 90° . Fractions at the bottom of bars are the number of connected over unconnected pairs in each bin. **c**, Synaptically connected neurons have greater potential synapse length. Cumulative distribution of potential synapse length was significantly greater between connected (red line) than unconnected pairs (blue line, $P \approx 0$, permutation test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$, $s = 5 \mu\text{m}$). Inset, schematic of a cylinder of length L_d (transparent purple) around the dendrite (red) with a radius (s) equivalent to a maximal spine's length where the axon (blue) comes within proximity to make a synapse. **d**, Cumulative distribution of differences in orientation preference was significantly less between connected (red line) than unconnected pairs (blue line, $P < 0.05$, permutation test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$) or a model distribution based on potential synapse length (black line, $P < 0.05$, permutation test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$). **e**, Potential synapse length (L_d) is uniform across differences in peak orientation preference between neurons in the model distribution. **f**, Synapse rate (λ , reconstructed synapses normalized by L_d) decreases with orientation preference difference across the reconstructed population ($P < 0.05$, Cochran–Armitage test, $n_{\text{synapses}} = 39$, $s = 5 \mu\text{m}$). Error bars in b, f and shaded regions in c, d represent bootstrapped standard error of the mean. Scale bar, a, $100 \mu\text{m}$.

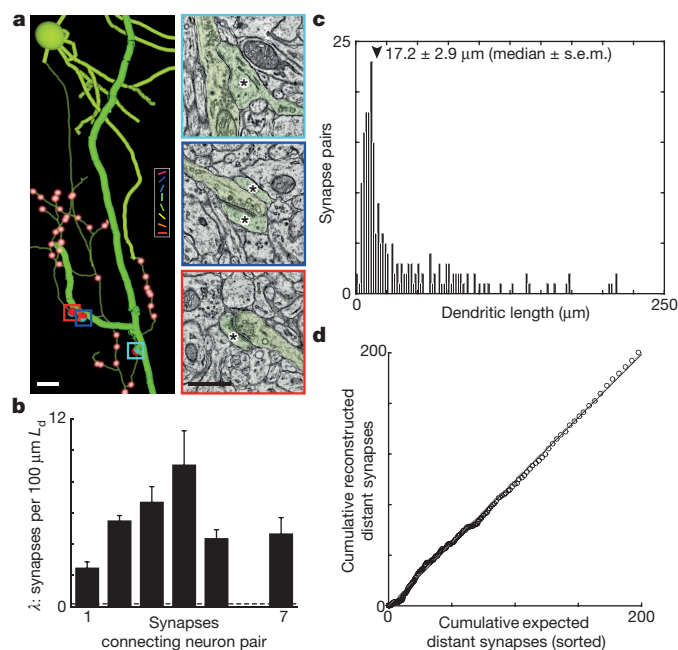


Figure 3 | Multiple synapses between neurons are found above chance levels; clustered synapses are frequently observed, but are not specifically enriched. **a**, Anatomical reconstruction (left) of a L2/3 pyramidal cell synapsing onto a deep layer apical dendrite (thick process) with multiple synapses (right, corresponding coloured boxes). Asterisks denote postsynaptic cell. Colour key and synapses not connecting this cell pair are rendered as in Figs 1 and 2. Of 130 reconstructed presynaptic neurons, 51 had axons making multiple synapses onto individual neuronal targets. **b**, Synapse rates (λ) normalized by potential synapse length (L_d , $s = 5 \mu\text{m}$) for neuron pairs connected by 1–7 synapses are significantly higher than expected from a Poisson process with a synapse rate $= \lambda$ ($P \approx 0$, permutation tests, $n_{\text{synapses}} = 292$). Dashed line is λ_{avg} from Fig. 2f. **c**, Histogram of the dendritic path length between pairs of synapses connecting neurons with multiple synapses. The median distance between synapses connecting neuron pairs was $17.2 \pm 2.9 \mu\text{m}$ (median \pm s.e.m., arrowhead). **d**, Multiple synapses are distributed randomly, with equal tendency to be clustered or distant. The cumulative number of distant ($>20 \mu\text{m}$) reconstructed synapses, vs. nearby ($<20 \mu\text{m}$), follows the cumulative number expected from a constant synapse rate, based on synapse multiplicity normalized by L_d , for each cell pair. x axis sorted by expected number of distant synapses; the dashed line is unity. Error bars, **b**, represent standard error. Scale bars, **a**, left, 10 and right, $1 \mu\text{m}$.

$>20 \mu\text{m}$ versus $<20 \mu\text{m}$ from each synapse between a pair of neurons, we found that the number of distant synapses correlated ($P \approx 0$, permutation test, $n_{\text{synapses}} = 195$, Pearson's $r = 0.92$) nearly perfectly with the expected value from the pair's average density of synapses (Fig. 3d, see Methods). That is, axons and dendrites of connected neurons generate multiple synapses when their axon and dendrite remain close, but make additional synapses with an equal rate (λ) elsewhere their processes overlap within our limited volume. This suggests plasticity mechanisms that operate at the cellular rather than dendritic level, but nonetheless leaves open a potential computational role for positioning multiple synapses, which occur both clustered and distributed along the dendritic arbor.

Finally, we examined synapse size by measuring postsynaptic density (PSD) area (Fig. 4a), which is proportional to the number of pre-synaptic vesicles and spine volume²⁶, and may be related to synaptic strength^{27–29}. We found that PSD areas could be large ($>0.25 \mu\text{m}^2$) for synapses between cells with similar peak orientations (Fig. 4b, c), but were clustered at a smaller size ($0.06 \pm 0.008 \mu\text{m}^2$, mean \pm s.e.m.) for the most dissimilar orientations. Connected cells (Fig. 4c, red line) with similar peak orientation preference exhibited larger PSD areas compared to control populations where the orientation selectivity

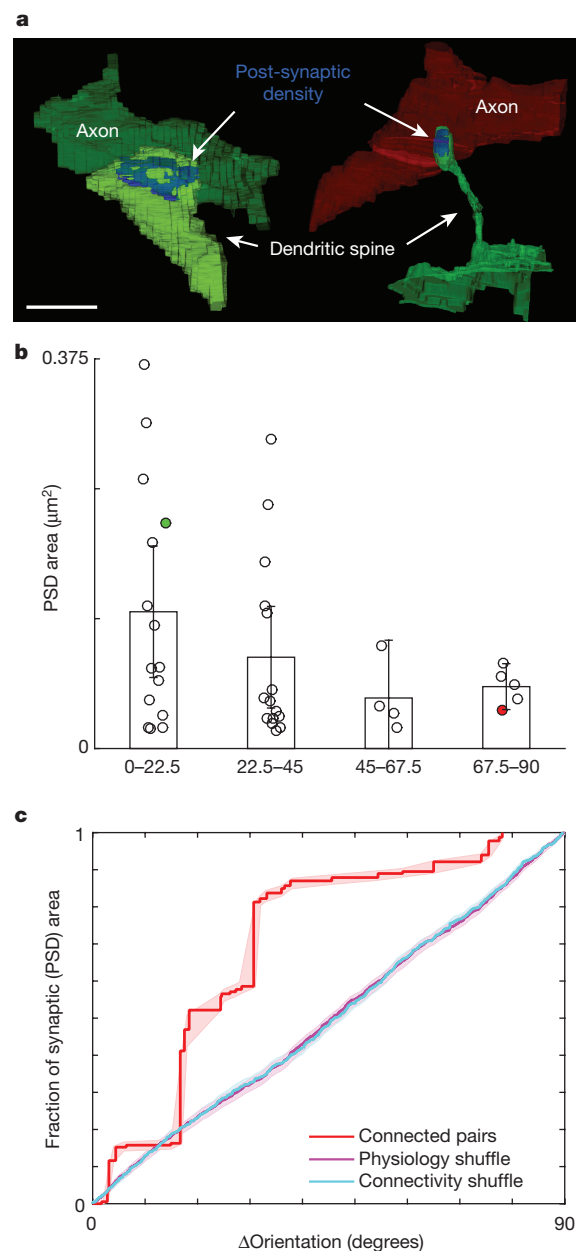


Figure 4 | Sensory physiology predicts synaptic size. **a**, Volumetric reconstructions of synapses between functionally characterized cells. Examples of postsynaptic densities (PSDs, blue) for a pair of iso-oriented neurons and for cross-oriented neurons (colour key as in previous figures). **b**, PSD area decreases as difference in orientation preference increases (examples in **a** labelled green and red, respectively; bins: 0° to 22.5° , 22.5° to 45° , 45° to 67.5° , and 67.5° to 90°). **c**, Cumulative fraction of PSD area accounted for by connections as a function of the difference in orientation preference (red line) is significantly greater than controls (shuffled orientation preference (magenta), or connectivity (cyan), both $P < 0.01$, permutation tests, $n_{\text{synapses}} = 39$). Error bars, **b**, represent 95% confidence intervals and shaded regions, **c**, represent bootstrapped standard error. Scale bar, **a**, $0.5 \mu\text{m}$.

differences or identity of connected cells were shuffled (Fig. 4c, cyan and magenta lines, respectively, $P < 0.01$, permutation tests). This functionally specific synaptic weighting was significant whether or not spines from apical dendrites were included in the analysis (data not shown). One possibility is that there are multiple classes of synapses, latent connections with smaller synapses, and larger synapses between similarly tuned cells.

Functional connectomics promises to build bridges between the *in vivo* activity of neurons, network connectivity, and neuronal

structure³⁰. Here, we used this approach to demonstrate five aspects of the excitatory cortical circuit. Independent of physiology, we demonstrate (1) a modular network organization that previously had only been inferred indirectly from smaller cortical subnetworks^{2,5,18,19}, and (2) that multiple synapses between pairs of neurons occur far above chance levels, often closely spaced on the dendrite ($<20\mu\text{m}$), although there seems to be no specific mechanism favouring local clustering over widespread spacing. Further, we demonstrate (3) an anatomical substrate of functionally specific connections between neurons, and (4) that this specificity does not result from the spatial arrangement of the neuropil, but instead must operate at the scale of dendritic spines ($<1\text{--}5\mu\text{m}$). Finally, we show (5) that synapse size correlates with physiology, with larger synapses found between neurons with similar peak orientations. Such specific connectivity is consistent with intracortical amplification of afferent signals, overcoming the strong inhibitory tone in the awake cortex^{3,4,15}. As methods for automated reconstruction improve, each of these findings will come into greater focus, leading to a richer understanding of network structure and function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 September 2014; accepted 2 February 2016.

Published online 28 March 2016.

- Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51–58 (2013).
- Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* (2011).
- Li, Y. T., Ibrahim, L. A., Liu, B. H., Zhang, L. I. & Tao, H. W. Linear transformation of thalamocortical input by intracortical excitation. *Nature Neurosci.* **16**, 1324–1330 (2013).
- Lien, A. D. & Scanziani, M. Tuned thalamic excitation is amplified by visual cortical circuits. *Nature Neurosci.* **16**, 1315–1323 (2013).
- Wertz, A. *et al.* Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science* **349**, 70–74 (2015).
- Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
- Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858–871 (2010).
- Bock, D. D. *et al.* Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).
- Hofer, S. B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neurosci.* **14**, 1045–1052 (2011).
- Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional specialization of mouse higher visual cortical areas. *Neuron* **72**, 1025–1039 (2011).
- Marshall, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72**, 1040–1054 (2011).
- Bonin, V., Histed, M. H., Yurgenson, S. & Reid, R. C. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *J. Neurosci.* **31**, 18506–18521 (2011).
- Glickfeld, L. L., Andermann, M. L., Bonin, V. & Reid, R. C. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nature Neurosci.* (2013).
- Cossell, L. *et al.* Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518**, 399–403 (2015).
- Reinhold, K., Lien, A. D. & Scanziani, M. Distinct recurrent versus afferent dynamics in cortical visual processing. *Nature Neurosci.* **18**, 1789–1797 (2015).
- Saalfeld, S., Cardona, A., Hartenstein, V. & Tomancak, P. CATMAID: collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* **25**, 1984–1986 (2009).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
- Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
- Yoshimura, Y., Dantzker, J. L. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).

- Bopp, R., Macarico da Costa, N., Kampa, B. M., Martin, K. A. & Roth, M. M. Pyramidal cells make specific connections onto smooth (GABAergic) neurons in mouse visual cortex. *PLoS Biol.* **12**, e1001932 (2014).
- Stepanyants, A., Hof, P. R. & Chklovskii, D. B. Geometry and structural plasticity of synaptic connectivity. *Neuron* **34**, 275–288 (2002).
- Kampa, B. M., Letzkus, J. J. & Stuart, G. J. Cortical feed-forward networks for binding different streams of sensory information. *Nature Neurosci.* **9**, 1472–1473 (2006).
- Reimann, M. W., King, J. G., Muller, E. B., Ramaswamy, S. & Markram, H. An algorithm to predict the connectome of neural microcircuits. *Front. Comput. Neurosci.* **9**, 120 (2015).
- Kasthuri, N. *et al.* Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).
- Bartol, T. M. *et al.* Nanoconnectomic upper bound on the variability of synaptic plasticity. *Elife* **4**, (2015).
- Harris, K. M. & Stevens, J. K. Dendritic spines of CA 1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* **9**, 2982–2997 (1989).
- Matsuzaki, M. *et al.* Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **4**, 1086–1092 (2001).
- Nusser, Z. *et al.* Cell type and pathway dependence of synaptic AMPA receptor number and variability in the hippocampus. *Neuron* **21**, 545–559 (1998).
- Takumi, Y., Ramirez-Leon, V., Laake, P., Rinvik, E. & Ottersen, O. P. Different modes of expression of AMPA and NMDA receptors in hippocampal synapses. *Nature Neurosci.* **2**, 618–624 (1999).
- Reid, R. C. From functional architecture to functional connectomics. *Neuron* **75**, 209–217 (2012).
- Rubinov, M. & Sporns, O. Weight-conserving characterization of complex functional brain networks. *Neuroimage* **56**, 2068–2079 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Arora, E. Ashbolt, L. Bailey, L. Beaton, P. Chopra, A. Coda, M. Driesbach, R. Fang, M. Fisher, A. Giuliano, H. Godtfredsen, A. Henry, E. Holtzman, P. Hughes, K. Joines, J. Larson, S. Maillet, K. Moody, E. Nguyen, A. Orban, V. Osuji, K. Robertson, D. Sandman, S. Schwartz, E. Szczerzenie, N. Thatra, L. Thomas, B. Titus, R. Torres for tracing and reconstruction; E. Raviola for discussions and advice; H. Kim for technical support at the beginning of the study; A. Wetzel for help with alignment and stitching; A. Cardona and S. Saalfeld for the CATMAID project being openly available; and N. da Costa for launching the EM annotation program at the Allen Institute for Brain Science, with help and support from B. Youngstrom, R. Young, C. Dang and J. Phillips. We also thank D. Brittain, W. Gray Roncal, L. Thomas, S. Yurgenson, H. Elliot, and the HMS Image and Data Analysis Core for programming; E. Benecchi and the HMS EM Core Facility for technical support; P.J. Manavalan, K. Lillaney, and R. Burns, for helping make the data freely available; and S. Druckmann, Y. Park, C. Priebe, and J. Vogelstein for discussions on statistical analyses. We are indebted to M. Andermann, S. Chatterjee, N. da Costa, L. Glickfeld, C. Harwell, D. Hildebrand, M. Histed, S. Mihalas, L. Ostroff, E. Raviola, and J. T. Vogelstein for critical reading of various versions of the manuscript. This work was supported by NIH grants to RCR (R01 EY10115 and R01 NS075436); through resources provided by the NRBSC (P41 RR06009) and MMBioS (P41 GM103712); the HMS Vision Core Grant (P30 EY12196); and the AIBS. We thank the Allen Institute founders, P. G. Allen and J. Allen, for their vision, encouragement, and support. WCL was also supported by the Bertarelli Program in Translational Neuroscience and Neuroengineering, Edward R. and Anne G. Lefler Center, and Stanley and Theodora Feldberg Fund, and the NIH (R21 NS085320). VB was also supported by Neuro-Electronics Research Flanders. The project described was partially supported by the NIH by the above named awards. Its content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author Contributions W.-C.A.L. and V.B. performed the *in vivo* calcium imaging and analysed it. W.-C.A.L. processed the tissue for EM, sectioned the series, and aligned the block with the *in vivo* imaging. W.-C.A.L. and M.R. imaged it on the TEMCA. G.H. and W.-C.A.L. aligned the EM images into a volume. W.-C.A.L., M.R., K.G. annotated the EM dataset and W.-C.A.L. and K.G. supervised segmentation efforts. B.J.G. and W.-C.A.L. generated software for visualization and analysis. W.-C.A.L. performed quantitative analysis on the tracing. W.-C.A.L., V.B., and R.C.R. designed the experiment and wrote the paper.

Author Information The aligned EM dataset will be publicly accessible at <http://neurodata.io/lee16>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.-C.A.L. (wei-chung_lee@hms.harvard.edu) or R.C.R. (clayr@alleninstitute.org).

METHODS

We imaged the primary visual cortex of an awake 9-month-old C57BL/6 male mouse, as described previously^{10,13}, with a custom-built two-photon microscope¹². Using volumetric *in vivo* two-photon calcium imaging of a genetically encoded calcium indicator (GCaMP3), we measured the time-resolved responses of a population of identified neurons to a wide array of stimuli including drifting gratings (up to 16 directions, 3 spatial, and 2 temporal frequencies). Following 12 days of imaging calcium responses in the same cohort of neurons, we labelled blood vessels with a tail vein injection (rhodamine B-conjugated dextran) and acquired an *in vivo* fluorescence volume. The animal's brain was then prepared for large-scale transmission EM as described previously⁸. 3,700 serial sections (<50 nm thick) were cut and imaged spanning a $450\mu\text{m} \times 450\mu\text{m} \times 150\mu\text{m}$ volume at $4\text{ nm} \times 4\text{ nm} \times 40\text{ nm}$ per voxel resolution. Sections representing the middle third of the EM volume were aligned and imported into CATMAID¹⁶ for distributed, online, manual reconstruction and targeted volumes around identified synapses were exported for volumetric segmentation and PSD analysis. EM reconstructed neurons were identified in the *in vivo* stack by using the blood vessels as landmarks. Apical dendrites originating from deeper neocortical lamina were similarly identified and corresponded by location and branching geometry of their apical tufts. Permutation tests were used in statistical analyses, unless otherwise noted.

Animal preparation. All procedures were conducted in accordance with the ethical guidelines of the NIH and approved by the IACUC at Harvard Medical School. For cranial window implant surgery the mouse was anesthetized with isoflurane (1.2–2% in 100% O₂). Dexamethasone (3.2 mg per kg body weight, intramuscular) was administered on the day before surgery and atropine (0.2 mg per kg body weight, intraperitoneally) at the beginning of surgery. Using aseptic technique, we secured a headpost in place using cyanoacrylate, dental acrylic, and C&B Metabond (Parkell), and made a 5 mm craniotomy over the left visual cortex (centre: ~2.8 mm lateral, 0.5 mm anterior to lambda) as described previously³². A 5 mm glass cranial window was implanted consisting of an 8 mm coverslip cured to two 5 mm coverslips (Warner #1; total thickness: ~0.5 mm; thickness below skull: ~200 µm) using index-matched adhesive (Norland #71). We secured the window in place using cyanoacrylate and dental acrylic.

We habituated the mouse with water scheduling so that water was delivered only during and immediately after head restraint training. We increased the duration of head restraint sessions over the course of 2 weeks, from 3 min to 2 h³². We then performed retinotopic mapping of visual cortical areas using widefield intrinsic autofluorescence imaging, measuring autofluorescence produced by blue excitation (470 nm centre, 40 nm band, Chroma) through a green/red emission filter (long-pass, 500 nm cutoff). We collected images using a CCD camera (Sensicam, Cooke, 344×260 pixels spanning $4\text{ mm} \times 3\text{ mm}$; 2 Hz acquisition rate) through a $5\times$ air objective (0.14 NA, Mitutoyo) using ImageJ acquisition software. For retinotopic mapping, stimuli were presented at 2–6 retinotopic positions for 10 s each, with 10 s of mean luminance between trials.

GCaMP3 expression was targeted by viral injection. Dexamethasone (3.2 mg per kg body weight, intramuscular) was administered at least 2 h before coverslip removal. The mouse was anesthetized (isoflurane, 1–1.5%) and the cranial window was sterilized with alcohol and the coverslip removed. We then volume injected ($50\text{--}100\text{ ml min}^{-1}$, Stoelting) $30\text{--}100\text{ nl}$ of a 10:1 mixture of AAV2/1. hSynap.GCaMP3.3.SV40³³ (Penn Vector Core) and 1 mM sulforhodamine-101 (Invitrogen) to visualize the injection. Using the blood vessel pattern observed during widefield imaging as a guide, we made an injection in the posterior part of primary visual cortex at a depth of ~250 µm below the pial surface. After injection, a new cranial window was sealed in place and the mouse recovered.

Visual stimulation. A 120 Hz LCD monitor (Samsung 2233RZ, 2200) was calibrated at each temporal frequency using a spectrophotometer (Photoresearch PR-650). We confirmed waveforms were sinusoidal by measuring luminance fluctuations of a full-field sinusoidally modulated stimulus (using a photomultiplier tube, Hamamatsu). The monitor was positioned so that the stimulus patch was 21 cm from the contralateral eye. Local 40° Gabor-like circular patches (sigmoidal 10–90% falloff in 10°) containing either square-wave (for mapping retinotopy with widefield intrinsic autofluorescence and targeting GCaMP3 injections) or sine-wave (for mapping position of receptive fields with two-photon imaging) drifting gratings (80% contrast) were alternated with periods of uniform mean luminance (59 cd m^{-2}). In an effort to increase the population of responsive cells and explore receptive field parameters we presented gratings of varying directions at multiple spatial and temporal frequencies or at different positions in the visual field. We presented either 8 directions at 3 spatial frequencies (0.06, 0.12, and 0.24 cycles per degree (cpd)) and 2 temporal frequencies (2 and 8 Hz), 16 directions at 2 spatial frequencies (0.04 and 0.16 cpd) and 2 temporal frequencies (2 and 8 Hz), 8 directions at 6 positions, or 16 directions at 4 positions (45–115° eccentricity and –5–25° elevation), for a total of 64 stimulus types plus 10% blank trials. Stimuli were centred on the location eliciting maximum calcium responses

in the imaged field (monocular cortex), which most effectively drove responses in the population for experiments that did not vary stimulus position. All stimuli in a given protocol were presented in a pseudo-random order (sampling without replacement), and presented 3 times per volume experiment with 2–4 experiments per volume per day.

Two-photon calcium imaging. Imaging was performed with a custom-designed two-photon laser-scanning microscope¹². Excitation light from a Mai Tai HP DeepSee laser (Spectra-Physics) with dispersion compensation was directed into a modulator (Conoptics) and a beam expander (Edmund Optics). The expanded beam was raster scanned into the brain with a resonant (4 kHz, Electro-Optical Products) and a conventional galvanometer (Galvoline) (240 line frames, bidirectional, 31 Hz) through a 0.8 numerical aperture (NA) $16\times$ objective lens (Nikon). Emitted photons were directed through a green filter (centre: 542 nm; band: 50 nm; Semrock) onto GaAsP photomultipliers (no cooling, 50 µA protection circuit, H7422P-40MOD, Hamamatsu). The photomultiplier signals were amplified (DHPCA-100, Femto), and low-pass filtered (cutoff frequency = ~700 kHz). These and the mirror driver signals were acquired at 3.3 MHz using a multifunction data acquisition board (PCI-6115, National Instruments). Images were reconstructed in MATLAB (MathWorks) and continuously streamed onto a RAID array. Microscope control was also performed in MATLAB using an analogue output board (PCI-6711, National Instruments). The laser's dispersion compensation was adjusted to maximize collected fluorescence.

A piezoelectric objective translator on the microscope enabled imaging multiple $300 \times 300 \times 100\mu\text{m}$ volumes with 8 planes at 4 Hz separated by ~12.5 µm allowing us to capture the response properties of many cells through the depth of L2/3. The imaged field of view was $200\text{--}300\mu\text{m}$ on a side at resolution of $0.8\text{--}1.2\mu\text{m}$ per pixel (dwell-time ~2.7 µs). GCaMP3 was excited at 920 nm. Laser power was automatically adjusted as a function of imaging depth at the modulator with power exiting the objective ranging from 30–60 mW.

During imaging, the mouse was placed on 6-inch diameter foam ball that could spin noiselessly on an axel (Plasteel). We monitored trackball revolutions using a custom photodetector circuit and recorded eye movements using an IR-CCD camera (Sony xc-ei50; 30 Hz) and infrared illumination (720–2,750 nm bandpass filter, Edmund). Visual stimuli were presented for 4 s with 4 s of mean luminance between trials. Recording sessions were 2–6 h in duration.

Use of the genetically encoded calcium indicator GCaMP3, permitted recording from the same neurons over multiple days with the selectivity of calcium signals stable over several days of imaging (Extended Data Fig. 1)^{32,34,35}. Within this volume we obtained calcium signals from cell bodies of superficial layer (L2/3) neurons and large calibre apical dendrites that continued beyond the depth of our imaging volume and had branching morphologies consistent with deep layer pyramidal cells. These were likely from L5 neurons because of their large calibre, and because most L6 pyramidal cells do not project their apical dendrites more superficially than L4^{36,37}. The calcium signals from these deep layer apical dendrites stem from either forward^{38,39} or back-propagating action potentials⁴⁰, are consistent across days (Extended Data Fig. 1) and along the length of the deep layer apical dendritic trunks (Extended Data Fig. 2), and therefore most likely reflect the response properties of the soma. We relocated the cohort of neurons daily by using the vasculature's negative staining as fiducial landmarks. For the final *in vivo* imaging session, we injected the tail vein with a fluorescent dye to label blood vessels (rhodamine B isothiocyanate–Dextran (MW ~70k), 5% v/v, Sigma) and acquired a fluorescence stack to correspond the calcium-imaged neurons *in vivo* with their identities in the EM volume *ex vivo*⁸ (see below, and Extended Data Fig. 4).

EM material preparation. Following *in vivo* two-photon imaging the animal was perfused transcardially (2% formaldehyde/2.5% glutaraldehyde in 0.1 M cacodylate buffer with 0.04% CaCl₂) and the brain was processed for serial-section TEM. $200\mu\text{m}$ thick coronal vibratome sections were cut, post-fixed, and *en bloc* stained with 1% osmium tetroxide/1.5% potassium ferrocyanide followed by 1% uranyl acetate, dehydrated with a graded ethanol series, and embedded in resin (TAAB 812 Epon, Canemco). We located the calcium-imaged region by matching vasculature between *in vivo* fluorescence and serial thick (1 µm) toluidine blue (EMS) sections cut from an adjacent vibratome sections, then cut ~3,700 serial (<50 nm) sections on an ultramicrotome (Leica UC7) using a 35 degree diamond knife (EMS-Diatome) and manually collected sections on 1 mm \times 2 mm dot-notch slot grids (Synaptek) that were coated with a pale gold Piloform (Ted Pella) support film, carbon coated, and glow-discharged. Following section pickup, we post-stained grids with uranyl acetate and lead citrate.

Large-scale TEM imaging, alignment, and correspondence. Using the custom-built transmission electron microscope camera array (TEMCA)⁸ we imaged the ~3,700 section series, targeting a $\sim 450\mu\text{m} \times 450\mu\text{m}$ region for each section (Fig. 1c). Acquired at 4 nm per pixel in plane, this amounted to ~100 terabytes of raw data to date comprising 30 million cubic microns of brain and >10 million ($4,000 \times 2,672$ pixel) camera images. Magnification at the scope was $2,000\times$, accelerating potential was

120 kV, and beam current was ~90 microamperes through a tungsten filament. Images suitable for circuit reconstruction were acquired at a net rate of 5–8 million pixels s^{-1} . Approximately the middle third of the series (sections 2,281–3,154) was aligned using open source software developed at Pittsburgh Supercomputing Center (AlignTK)⁸ and imported into CATMAID¹⁶ for distributed online visualization and segmentation. Within the analysed EM series there were 51 missing sections. Nineteen were single section losses. There were 2 instances each of missing 2, 3, and 4; and 1 instance each of missing 6 or 8 consecutive sections near the series boundaries. Folds, staining artefacts, and sometimes cracks occurred during section processing, but were typically isolated to edges of our large sections and therefore did not usually interfere with manual segmentation. To find the correspondence between the cells imaged *in vivo* with those in the EM data set, a global 3D affine alignment was used with fiducial landmarks manually specified at successively finer scales of vasculature and then cell bodies to re-locate the calcium-imaged neurons in the EM-imaged volume (Extended Data Fig. 4). Apical dendrites arising from deep layer (putative L5) pyramidal neurons were identified by their characteristic morphology^{36,41,42} (also see below). Their correspondence was facilitated by the unique branching patterns of their apical tufts and those that could not be unambiguously identified were not included in the functional analysis.

Reconstruction and verification. We first traced the axonal and dendritic arbors of the functionally characterized neurons in the EM data set by manually placing a series of marker points down the midline of each process to generate a skeletonized model of the arbors using CATMAID¹⁶ (Figs 1d, 2a, 3a, Extended Data Fig. 6, Supplementary Data 1–3). We identified synapses using classical criteria⁴². For each synapse on the axon of a functionally characterized cell, dendrites of postsynaptic excitatory neurons were traced either to the boundaries of the aligned volume or centripetally back to the cell body⁸. We identified deep layer apical dendrites of (putative L5) pyramidal cells by their large calibre, high spine density, and their continuation beyond the bottom border of the EM volume, which spans from the pial surface through L4. For each neuronal target reconstruction included in the analysis, a second independent annotator verified the tracing by working backwards from the most distal end of every process. An additional round of validation was done for each synapse between functionally characterized cells where a third annotator who had not previously traced the pre- or post-synaptic process, independently verified the anatomical connectivity blind to previous tracing work. We began this independent round of validation at each synapse and traced the pre- and postsynaptic processes centripetally. If the initial reconstruction and subsequent verification of the reconstruction diverged, that connection and the segmentation work distal from the point of divergence was excluded from further analysis. EM reconstruction and validation was performed blind to cells' functional characteristics and targeted cells were initially assigned to individual annotators pseudo-randomly weighted by tracing productivity.

We performed targeted volumetric reconstructions of synapses connecting functionally characterized cells by developing tools to interface with CATMAID cutout, locally align, and catalogue volumes of interest based on location (Fig. 4a; for example, 400 pixels \times 400 pixels \times 41 sections or 3.2 $\mu m \times$ 3.2 $\mu m \times$ 1.64 μm volumes centred on synapses represented by CATMAID connectors). Presynaptic boutons, postsynaptic spines, their parent axons and dendrites, and postsynaptic density (PSD) areas were manually segmented with itk-SNAP (<http://www.itksnap.org/>). PSD areas were calculated as described previously⁴³ with obliquely cut or *en face* synapse areas measured using their maximum *z*-projection. *En face* or obliquely cut synapses were identified by serial sections that starkly transitioned from a clear presynaptic specialization hosting a vesicle pool, to a distinctly different postsynaptic cell, typically with an intervening section of electron dense area representing the postsynaptic density and/or synaptic cleft (for example, Extended Data Fig. 5).

Data analysis. *In vivo* calcium imaging data was analysed in MATLAB and ImageJ (NIH) as described previously^{12,13}. To correct for motion along the imaged plane (*x*–*y* motion), the stack for each imaging plane was registered to the average field of view using TurboReg⁴⁴. A 5 pixel border at each edge of the field of view was ignored to eliminate registration artefacts. Masks for analysing fluorescence signal from neurons were manually generated corresponding to cells in the EM volume, registered to the *in vivo* anatomical fluorescence stack, and to individual physiological imaging planes. Time courses of cells spanning multiple physiological imaging planes were weighted by dwell time in each plane and averaged across planes. Evoked responses for each EM identified cells were measured for each stimulation epoch as the difference in fractional fluorescence (% $\Delta F/F_0$) between the 5 s after and the 2.5 s before stimulus onset (pre-stimulus activity), and averaged across stimulus repetitions. We quantified visual responsiveness of each cell by calculating the average Pearson correlation coefficient of the responses to all stimuli across repetitions (average trial-to-trial correlation). We defined the significance of visual responses as the probability (*P* value) that the observed trial-to-trial correlation is larger than the correlation obtained from a full random permutation of the data for

spatial and temporal frequency experiments ($P_{\text{sfr}} < 0.05$) and experiments where stimulus position was varied ($P_{\text{pos}} < 0.01$). In retinotopic experiments designed to increase the number of characterized neurons, we found cells that did not reliably respond to stimuli $\pm 20^\circ$ from the centre of the display. These cells that either had receptive fields smaller than our stimuli or stimuli were positioned at the edge of their receptive fields. We considered these cells as potentially driven by stimulus edge effects and therefore excluded such experiments from further analysis. To estimate the preferred orientation, direction, and spatiotemporal frequency, we modelled responses with a combination of a multivariate Gaussian with spatial frequency (*x* and *y*, deg), temporal frequency (Hz) and position (*x* and *y*, deg) as independent dimensions, a constant gain factor, and a static exponent. We fit the model to data using a large-scale nonlinear optimization algorithm (Trust Region Reflective, MATLAB Optimization Toolbox, MathWorks Inc.), generating multiple fits from randomly selected starting points and selected the best fit (least-square criterion). The quality of model fits was inspected visually for all neurons included in the data set.

EM connectivity was analysed using custom written software in MATLAB and Python. Connectivity analysis that did not utilize functional information (Figs 1e and 3, Extended Data Fig. 7) started with the entire population of excitatory neuronal targets in the reconstructed network. Network modularity and neuron connectivity motifs (Fig. 1e and Extended Data Fig. 7) were analysed with code modified from the Brain Connectivity Toolbox⁴⁵. We used an implementation of the Louvain method¹⁷ followed by consensus partitioning⁴⁶ for weighted and directed graphs to detect communities, or interconnected pyramidal neuron targets, from our EM reconstructed network purely by anatomical connectivity. For this analysis we included only the 201 traced neurons having multiple synaptic partners (degree ≥ 2). The number of synapses reconstructed between neurons was used as weights for all analyses. Modularity *Q* was given by the standard equation:

$$Q = \frac{1}{l} \sum_{i,j \in N} \left[a_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{l} \right] \delta(m_i, m_j)$$

where *l* is the total number of edges, given by

$$l = \sum_{i,j \in N} a_{ij}$$

where *N* is the total set of nodes, a_{ij} is the (*i*, *j*) element of the weighted adjacency matrix, $\delta(m_i, m_j)$ is 1 if *i* and *j* are in the same community and 0 otherwise, k_j^{in} and k_i^{out} are the in and out degrees of the *j*th and *i*th nodes respectively, calculated by

$$k_j^{\text{in}} = \sum_{i \in N} a_{ij} \quad k_i^{\text{out}} = \sum_{j \in N} a_{ij}$$

To generate null models of connectivity matrices for hypothesis testing, we shuffled the reconstructed adjacency conditioned on our sample degree, weight and strength distributions (Extended Data Fig. 7)^{31,47}.

Analysis of connectivity with neuronal function restricted our sample population to those cell pairs where both pre- and post-synaptic cells were functionally characterized. For orientation tuning (Figs 1d, f, 2, 4a–c, Extended Data Figs 5, 6, 8, 9), between 50 neurons, there were 29 connected pairs. On average, we detected 1.3 synapses per connected pair where we measured orientation selectivity for both cells. We varied retinotopic position and spatial and temporal frequencies of the grating stimulus with the goal of improved measurement of orientation preference for more cells. The sensory physiology of a subset of cells were simultaneously recorded across multiple stimulus parameters. These 120 cells were used for signal correlation analysis (Extended Data Fig. 10).

Potential synapse length (L_d) represents the degree to which pairs of neurons' axonal and dendritic arbors come sufficiently close to make a synapse (Fig. 2a, c–f, 3b, d, Extended Data Figs 9, 10). For excitatory pyramidal cells, we computed this length of potential synaptic connectivity between all pairs by first resampling the dendritic and axonal arbor skeletons to a maximum segment length of 40 nm (the average thickness of the EM sections) and summing the length of all dendrite segments within a maximum spine length distance of the axon ($s = 5 \mu m$; Figs 2, 3 and Extended Data Fig. 10; $s = 1 \mu m$: Extended Data Fig. 9). We use $s = 5 \mu m$ based the longest spine connecting functionally connected neurons ($\sim 5 \mu m$).

Analysis of neurons connected by multiple synapses (Fig. 3) was not restricted to cell pairs where both pre and post-synaptic cells were physiologically characterized. This population included 137 neurons connected by 267 synapses in 115 multi-synapse cell pairs whose axonal and dendritic arbors were traced exhaustively in the aligned volume. As a comparison population, we used 25 unique pairs connected by one synapse from the functionally characterized population described above, because they were also reconstructed throughout the aligned volume. To examine whether poly-synaptic connectivity occurs greater than random, we first computed a population average synapse rate (λ_{avg}) normalized by potential synapse

length, by dividing the total number of synapses reconstructed from the set of 50 functionally characterized neurons by their total pairwise L_d . We next compared λ for individual neuron pairs each connected by different numbers of synapses (Fig. 3b). This was used to assess whether multiple synapses occurred more often than predicted from a simple Poisson model.

We examined the frequency of clustered vs distant synapses by comparing synapse pairs that were separated by $>20\mu\text{m}$ or $<20\mu\text{m}$. For each synapse from each pair of neurons connected by n synapses, we computed the total L_d within $20\mu\text{m}$ or beyond $20\mu\text{m}$ from that synapse. We then took the fraction of the overlap beyond $20\mu\text{m}$:

$$L_d(>20\mu\text{m})/[L_d(<20\mu\text{m}) + L_d(>20\mu\text{m})]$$

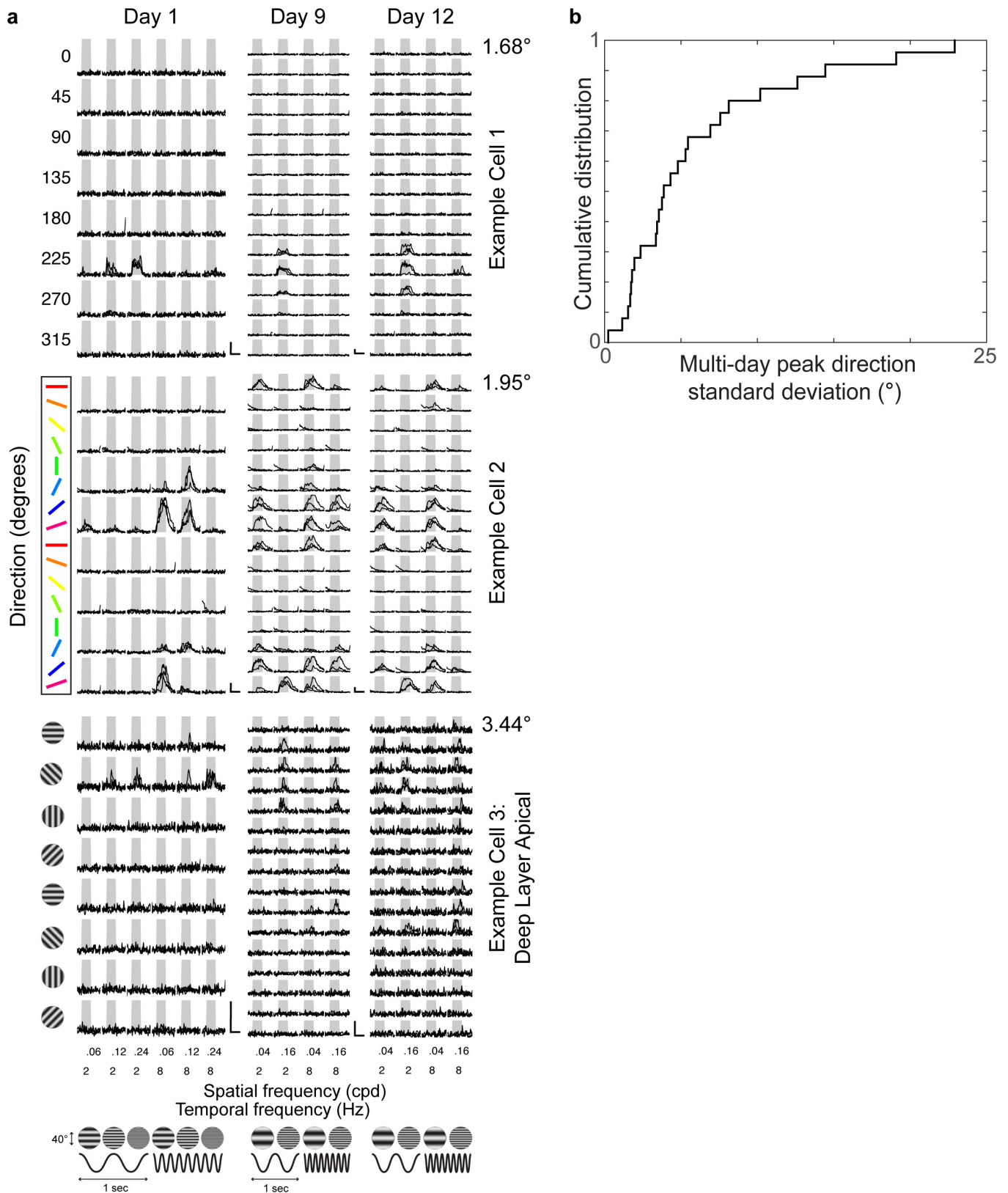
as the expected probability that each of the $(n - 1)$ other synapses will occur $>20\mu\text{m}$ away. The expected number of distant synapse was taken as $(n - 1)$ times the fraction of overlap beyond $20\mu\text{m}$, which was compared with the actual number of distant synapses observed (Fig. 3d).

3D renderings were generated using Blender (<http://www.blender.org/>) (Figs 1d, 2a, 3a, Extended Data Fig. 6, Supplementary Data 1–3), Imaris (Bitplane) (Extended Data Fig. 4 and Supplementary Video 1), and itk-SNAP (Fig. 4a). Cytoscape (<http://www.cytoscape.org/>) was used for network graph layouts (Figs 1f).

Statistics. Statistical methods were not used to predetermine sample sizes. Statistical comparisons between sample distributions were done with Permutation tests (that is, Monte Carlo-based Randomization tests) unless otherwise noted. Permutation tests were ideal as we do not assume the underlying distributions are normal, nor need the observations to be independent. For Permutation tests, we computed the incidence of differences between means or Pearson's linear correlation coefficient of randomly drawn samples from combined sample distributions exceeding the empirical difference (Figs 2b–d, f, 4c, and Extended Data Figs 7b, 9a, b, 10c, d). Cochran-Armitage two-sided tests for trend were used on proportional binned data with linear weights (Fig. 2b, f). Standard errors were calculated from bootstrapped sample distributions. For cumulative distributions (Figs 2c, d, 4c, and Extended Data Figs 9a, b, 10c, d), we repeatedly resampled by randomly drawing with replacement from the sample distribution the number of observed values 1,000–10,000 times and extracted the standard deviation at each step of the empirical CDF. For binned data (Fig. 2b, f, and Extended Data Fig. 9d), each resampled distribution was binned and the standard deviation was computed from the resampled probabilities or rates within each bin.

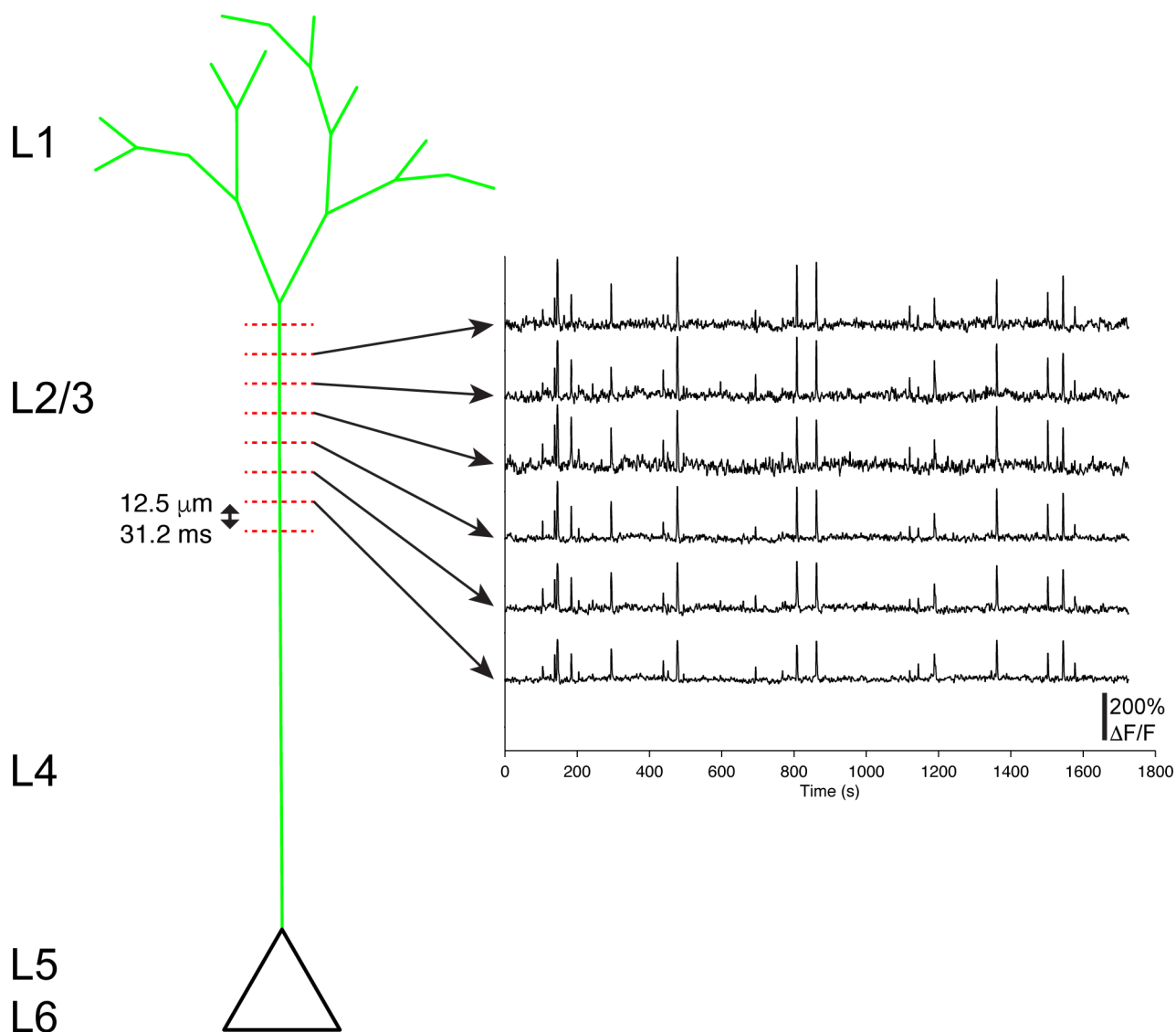
Code availability. Custom code is available upon request.

32. Andermann, M. L., Kerlin, A. M. & Reid, R. C. Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing. *Front. Cell. Neurosci.* **4**, 3 (2010).
33. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
34. Mank, M. *et al.* A genetically encoded calcium indicator for chronic *in vivo* two-photon imaging. *Nature Methods* **5**, 805–811 (2008).
35. Andermann, M. L. *et al.* Chronic cellular imaging of entire cortical columns in awake mice using microprisms. *Neuron* **80**, 900–913 (2013).
36. Peters, A. & Kara, D. A. The neuronal composition of area 17 of rat visual cortex. I. The pyramidal cells. *J. Comp. Neurol.* **234**, 218–241 (1985).
37. Ferrer, I., Fabregues, I. & Condom, E. A Golgi study of the sixth layer of the cerebral cortex. I. The lissencephalic brain of Rodentia, Lagomorpha, Insectivora and Chiroptera. *J. Anat.* **145**, 217–234 (1986).
38. Hirsch, J. A., Alonso, J. M. & Reid, R. C. Visually evoked calcium action potentials in cat striate cortex. *Nature* **378**, 612–616 (1995).
39. Smith, S. L., Smith, I. T., Branco, T. & Hausser, M. Dendritic spikes enhance stimulus selectivity in cortical neurons *in vivo*. *Nature* **503**, 115–120 (2013).
40. Markram, H., Helm, P. J. & Sakmann, B. Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. *J. Physiol. (Lond.)* **485**, 1–20 (1995).
41. Larkman, A. & Mason, A. Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. I. Establishment of cell classes. *J. Neurosci.* **10**, 1407–1414 (1990).
42. Peters, A., Palay, S. L. & Webster, H. d. *The fine structure of the nervous system: neurons and their supporting cells* 3rd edn (Oxford Univ. Press, 1991).
43. Harris, K. M. & Stevens, J. K. Dendritic spines of rat cerebellar Purkinje cells: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* **8**, 4455–4469 (1988).
44. Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).
45. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
46. Lancichinetti, A. & Fortunato, S. Consensus clustering in complex networks. *Sci. Rep.* **2**, 336 (2012).
47. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).

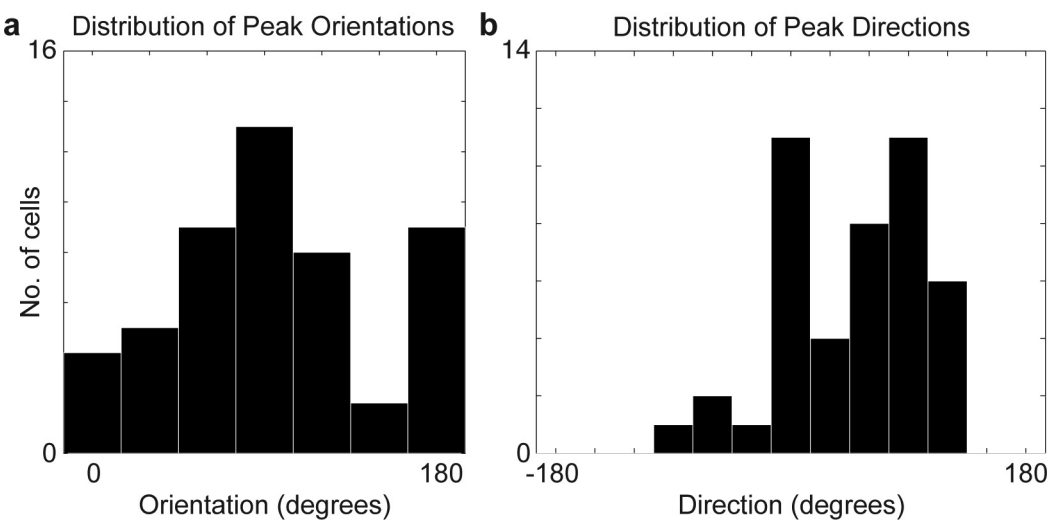


Extended Data Figure 1 | Sensory physiology is maintained over days. **a**, Visually evoked calcium responses are maintained across days. Three example neurons (rows) on the first, ninth, and twelfth day (columns) of *in vivo* two-photon calcium imaging. Three individual trial responses (black lines) of 8 directions, 3 spatial and 2 temporal frequencies (left column, day 1); or 16 directions, 2 spatial and 2 temporal frequencies

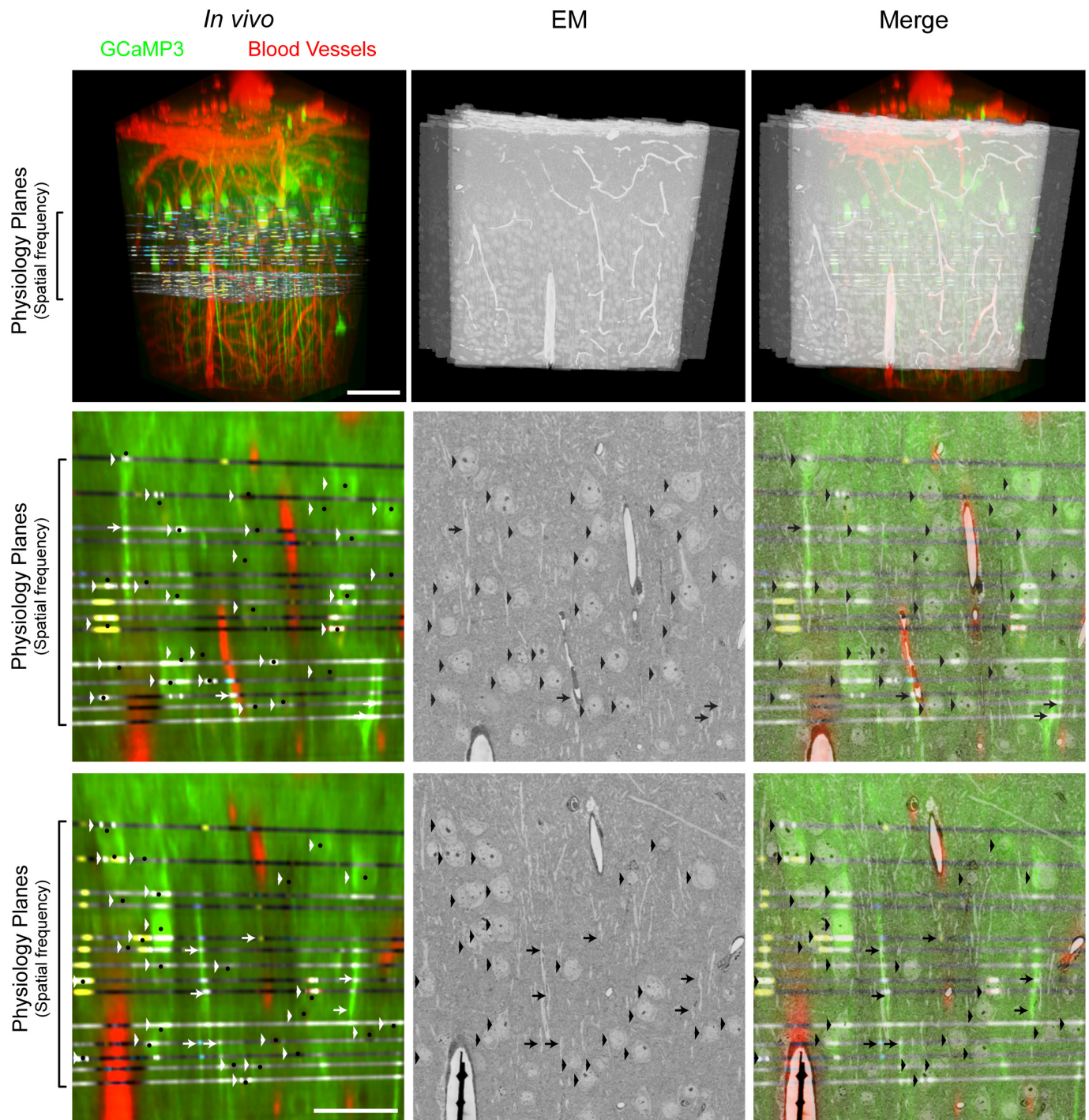
(middle and right rows, days 9 and 12). Scale bars, 100% $\Delta F/F$ and 4 s. Standard deviation of preferred direction across days for each neuron is to the upper right of activity matrices. **b**, Neurons direction selectivity is stable over days. Cumulative distribution of standard deviation of peak preferred direction ($4.1^\circ \pm 1.7^\circ$, median \pm s.e.m.) across days for 25 neurons measured over multiple days.



Extended Data Figure 2 | Deep layer apical responses are likely suprathreshold reflecting activity at the soma. Example $\Delta F/F$ time courses from a deep layer apical dendrite optically sectioned *in vivo* across 6–8 planes. Note, activity is correlated across depth and relatively stable over days (Extended Data Fig. 1).

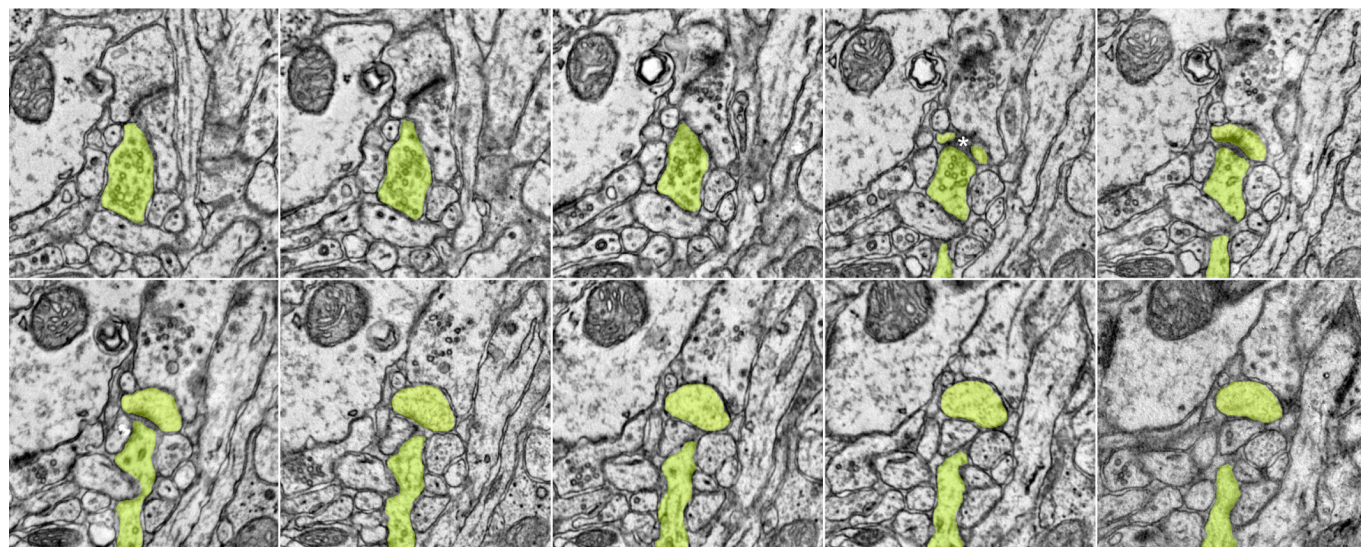


Extended Data Figure 3 | Distribution of orientation and direction tuned cells. **a, b,** Histograms of functionally characterized cells with measured peak orientation (**a**) and direction preference (**b**).

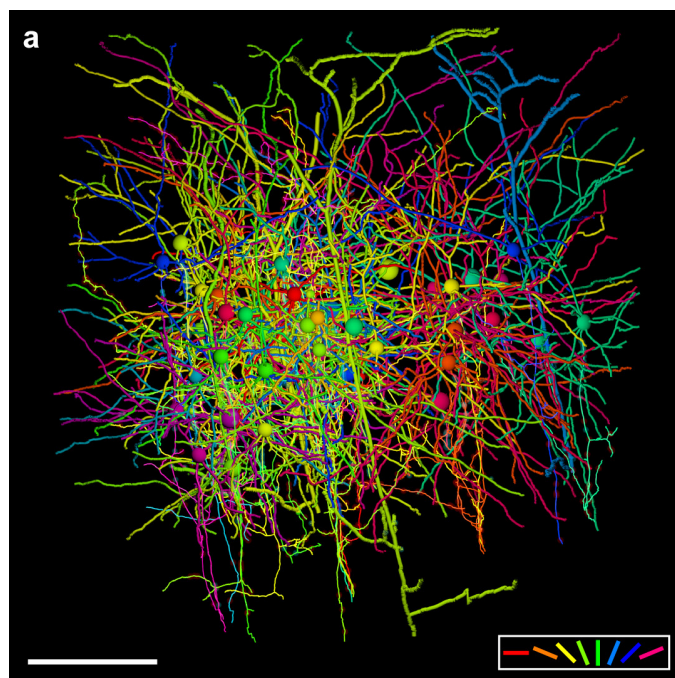


Extended Data Figure 4 | *In vivo* to EM correspondence of neuronal targets. Top row, volumetric projections of the aligned *in vivo* and EM imaged volumes. Physiology planes were acquired horizontally and EM sections cut frontally (coronally) from the brain. Interdigitated physiology planes are from 2 representative volumetrically scanned experiments stacked atop one another in space (Fig. 1c) so as to span from the border

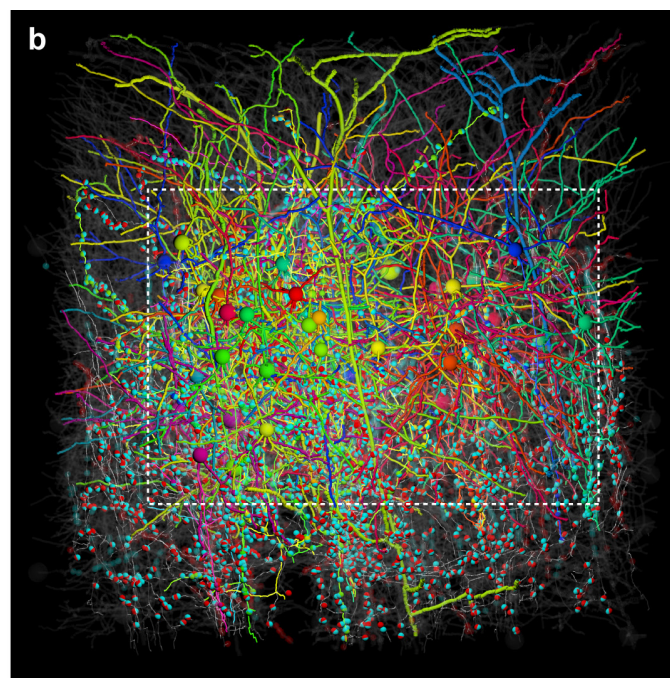
of L1 and L2 through the depth of L2/3. Scale bar, 100 μm . Middle and bottom rows, re-sliced planes through the *in vivo* volume corresponding to EM sections. Arrowheads indicate matching cell bodies, and arrows indicate deep layer (putative L5) apical dendrites. Small black dots mark the centres of cell bodies corresponding principally to nuclei where calcium indicator fluorescence is typically excluded. Scale bar, 50 μm .



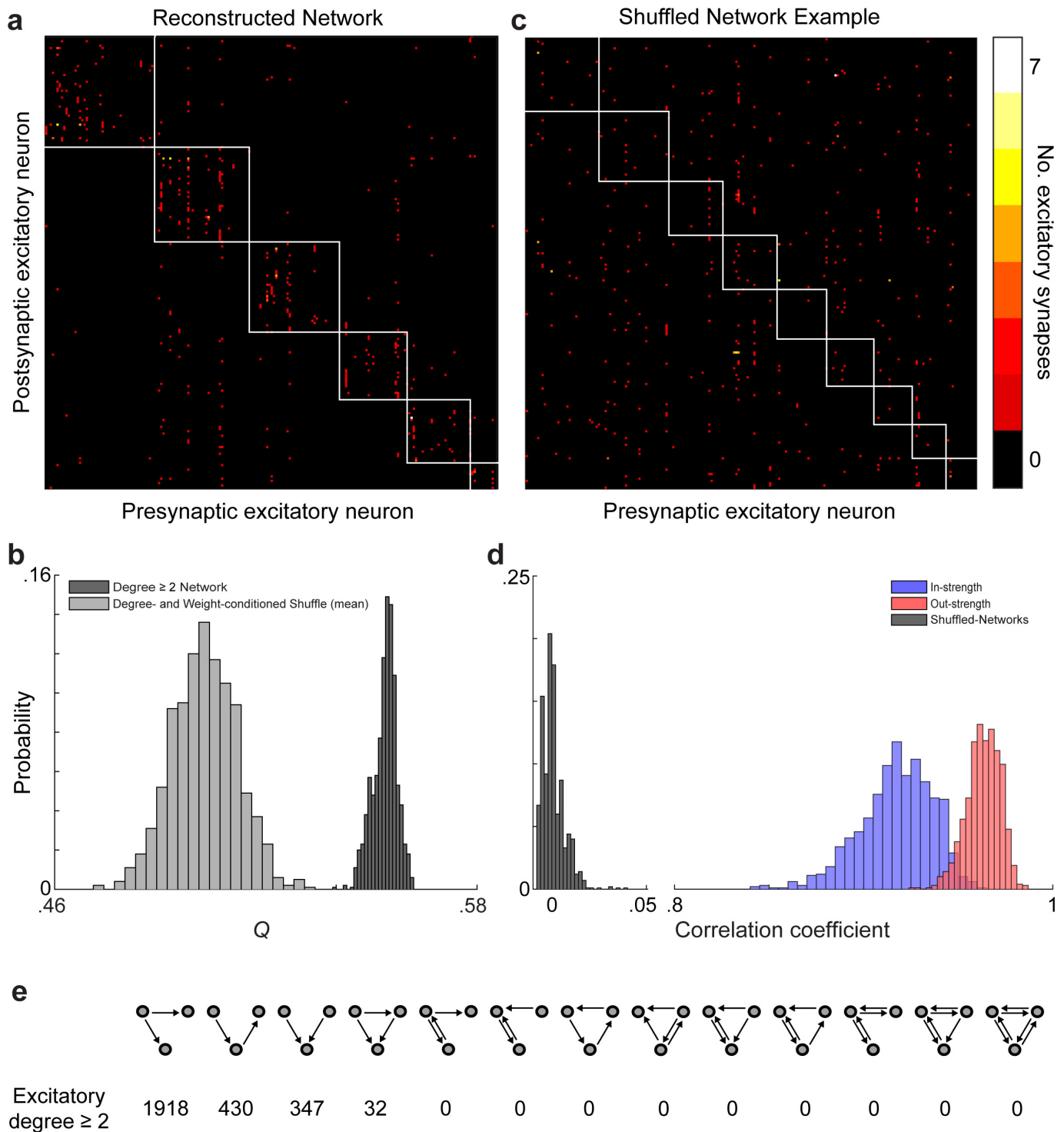
Extended Data Figure 5 | *En face* synapse example. Serial sections from an obliquely cut synapse (Fig. 1d, top right, same scale). Colour overlays correspond to peak orientation preference (colour key, same as Fig. 1a, bottom).



Extended Data Figure 6 | Network reconstruction. 3D rendering of dendrites and axons, cell bodies (large spheres), and synapses (small spheres) of, **a**, 50 functionally characterized neurons reconstructed in the EM volume. Cell bodies, dendrites, axons, and synapses colour coded by peak preferred stimulus orientation (colour key, bottom right). Axons are the thinnest processes, dendrites of L2/3 neurons are thicker, and deep

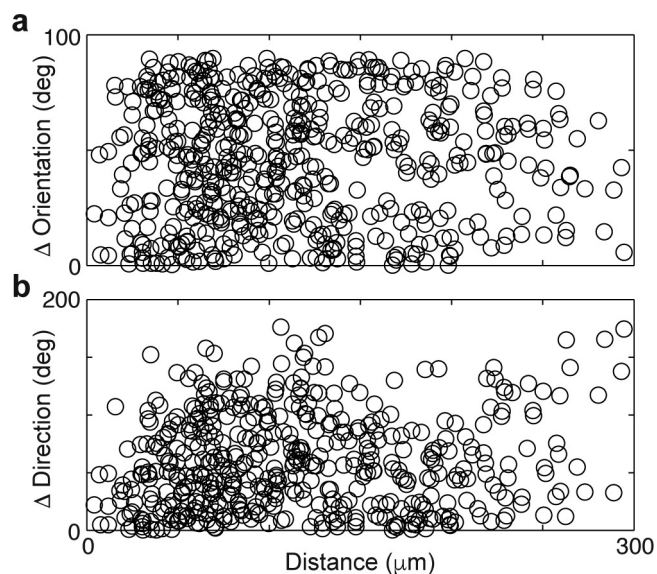


layer apical dendrites are rendered with the largest calibre. Dendritic spines were traced only if they participated in connections between reconstructed neurons. **b**, Approximately 1,800 additional neuronal targets reconstructed in the EM volume (transparent grey). Input and output synapses are coloured cyan and red respectively when orientation selectivity was not known. Bounding box matches region in Fig. 1f. Scale bar, 150 μm .

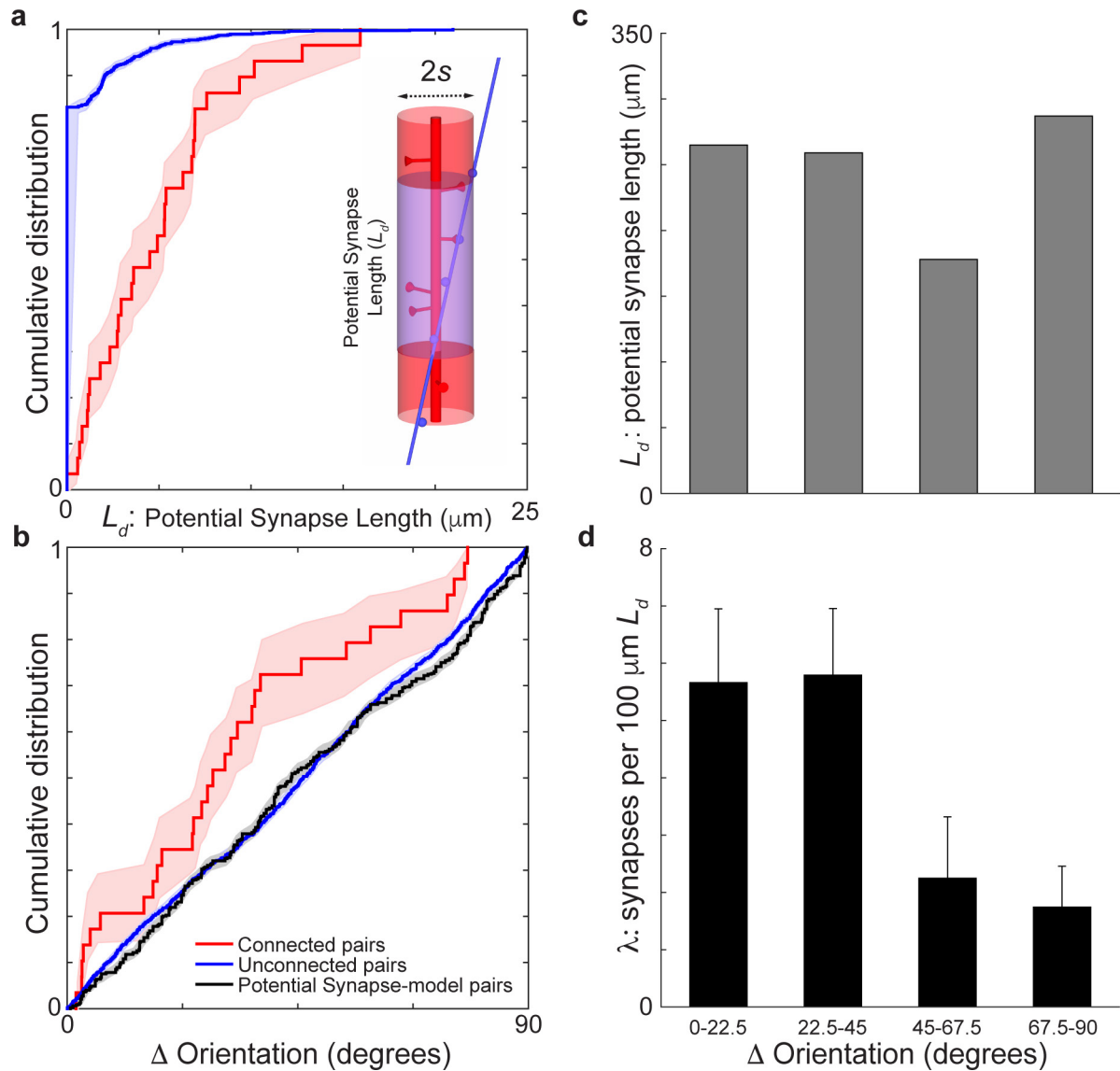


Extended Data Figure 7 | Network modularity is significantly non-random. **a**, Connectivity matrix of 201 excitatory neuronal targets in our network reconstruction with multiple synaptic partners (that is, degree ≥ 2 , no leaf nodes, same as Fig. 1e). Colour represents the number of synapses (colour key, **c**, right) between pre- and post-synaptic neurons (same neuron order on both ordinate and abscissa). Subnetworks of interconnected neurons (white boxes) detected using a consensus method of Louvain clustering^{17,31}. **b**, Modularity (Q) of the reconstructed network is significantly greater than null models with degree, weight, and strength preserved. Histograms of the modularity values for the reconstructed network (dark grey, $Q_{\text{mean}} = 0.55 \pm 0.003$, mean \pm s.d., computed 1,000 times)

is significantly greater than for the Q_{mean} of shuffled null models (light grey, $Q_{\text{mean}} = 0.50 \pm 0.009$, mean \pm s.d., $P \approx 0$, permutation test, $k_{\text{shuffles}} = 1,000$). **c**, Example of the shuffled connectivity matrix with a Q closest to the mean of the shuffled distribution with clustering (white boxes) computed as in **a**. **d**, Null models are well-shuffled, while approximating connection input and output strengths. Histograms of correlation coefficients between the reconstructed network and the null models' in (blue: 0.92 ± 0.02 , mean \pm s.d.) and out (red: 96 ± 0.01 , mean \pm s.d.) strength and connectivity matrix (grey: $9.1 \times 10^{-4} \pm 0.01$, mean \pm s.d.). **e**, Occurrences of three neuron connectivity motifs found in the reconstructed network between excitatory neuronal targets.

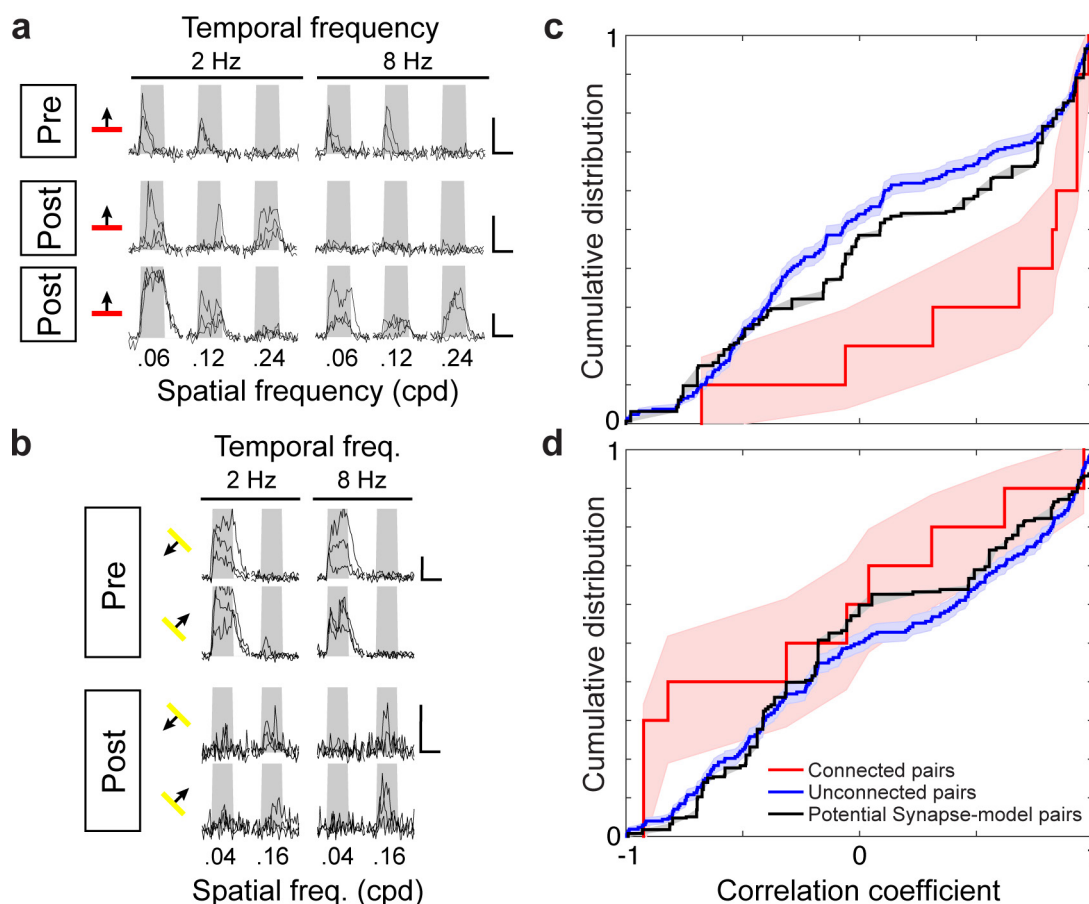


Extended Data Figure 8 | Cell bodies are functionally intermingled.
a, b, Differences in peak orientation (**a**) and direction preference (**b**) between neuron pairs plotted against the distance between their cell bodies. Uniform distributions of functional versus spatial distance suggest a salt and pepper intermingling of neuronal cell bodies across functional properties.



Extended Data Figure 9 | Axons and dendrites are functionally intermingled at shorter length scales. Uniform functional diversity and prediction of connectivity at finer length scales suggest a salt-and-pepper intermingling of axons and dendrites. **a–d**, Same as Fig. 2c–f for $s = 1 \mu\text{m}$. Significance tests: **a**, $P \approx 0$, permutation test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$;

b, Between connected (red line) and unconnected pairs (blue line, $P < 0.05$, permutation test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$) or a model distribution based on potential synapse length (black line, $P < 0.01$, permutation test, $n_{\text{connected pairs}} = 29$, $n_{\text{unconnected pairs}} = 1,951$). Shaded regions, **a**, **b**, and error bars, **d**, represent bootstrapped standard error.



Extended Data Figure 10 | Connectivity is not predicted by residual signal correlation after removal of orientation preference. **a–b**, Example activity ($\Delta F/F$ individual trial time courses) for connected neurons from experiments varying direction, spatial and temporal frequencies of grating stimuli. **a**, Presynaptic cell (top row) and two of its postsynaptic partners' (middle and bottom rows) for 3 spatial and 2 temporal frequencies and one orientation (orientation tuning was virtually identical). **b**, Presynaptic cell (top) and a postsynaptic deep layer apical dendrite's (bottom) responses to 2 spatial and 2 temporal frequencies, and 2 directions, (again, orientation tuning was virtually identical). Grey window delineates time of stimulus presentation. Scale bars, 100% $\Delta F/F$ and 4 s. **c**, Cumulative distribution of

signal correlations from simultaneously measured cells was significantly greater between connected than unconnected pairs ($P < 0.01$, permutation test, $n_{\text{connected pairs}} = 10$, $n_{\text{unconnected pairs}} = 426$) or a model distribution based on potential synaptic connectivity ($P < 0.05$, permutation test). **d**, After averaging over orientations, the cumulative distribution of signal correlations was similar between connected and unconnected pairs ($P > 0.14$, permutation test, $n_{\text{connected pairs}} = 10$, $n_{\text{unconnected pairs}} = 426$) and a model distribution based on potential synaptic connectivity ($P > 0.25$, permutation test, $n_{\text{connected pairs}} = 10$, $n_{\text{unconnected pairs}} = 426$). Shaded regions, **c**, **d**, represent bootstrapped standard error.

Daily magnesium fluxes regulate cellular timekeeping and energy balance

Kevin A. Feeney¹, Louise L. Hansen², Marrit Putker¹, Consuelo Olivares-Yañez³, Jason Day⁴, Lorna J. Eades⁵, Luis F. Larrondo³, Nathaniel P. Hoyle¹, John S. O'Neill¹ & Gerben van Ooijen²

Circadian clocks are fundamental to the biology of most eukaryotes, coordinating behaviour and physiology to resonate with the environmental cycle of day and night through complex networks of clock-controlled genes^{1–3}. A fundamental knowledge gap exists, however, between circadian gene expression cycles and the biochemical mechanisms that ultimately facilitate circadian regulation of cell biology^{4,5}. Here we report circadian rhythms in the intracellular concentration of magnesium ions, $[Mg^{2+}]_i$, which act as a cell-autonomous timekeeping component to determine key clock properties both in a human cell line and in a unicellular alga that diverged from each other more than 1 billion years ago⁶. Given the essential role of Mg^{2+} as a cofactor for ATP, a functional consequence of $[Mg^{2+}]_i$ oscillations is dynamic regulation of cellular energy expenditure over the daily cycle. Mechanistically, we find that these rhythms provide bilateral feedback linking rhythmic metabolism to clock-controlled gene expression. The global regulation of nucleotide triphosphate turnover by intracellular Mg^{2+} availability has potential to impact upon many of the cell's more than 600 Mg ATP-dependent enzymes⁷ and every cellular system where Mg NTP hydrolysis becomes rate limiting. Indeed, we find that circadian control of translation by mTOR⁸ is regulated through $[Mg^{2+}]_i$ oscillations. It will now be important to identify which additional biological processes are subject to this form of regulation in tissues of multicellular organisms such as plants and humans, in the context of health and disease.

Circadian rhythms occur cell-autonomously and are not restricted to metazoans or multicellular organisms, being found throughout eukaryotes and some prokaryotes⁹. Although explicit clock gene identities share no similarity across phylogenetic kingdoms, in every case temporal orchestration of gene expression is driven by timekeeping mechanisms that result in rhythmic clock protein transcription factor activity. In human cells, for example, a heterodimeric complex of BMAL1 and CLOCK positively regulates the expression of genes (period1/2, cryptochrome1/2) that encode its own repressor complex⁹, whereas in the marine unicellular alga *Ostreococcus tauri*, a reduced version of the stereotypical plant-like circadian clock consists of a feedback loop between the morning-expressed MYB-like transcription factor CCA1 and the evening-expressed protein TOC1¹⁰. Intriguingly, the role of enzymes such as casein kinase 1 and 2 in the post-translational regulation of clock protein activity/stability, and the speed at which biological clocks run, is functionally conserved across eukaryotes¹¹. Also, we recently reported a circadian rhythm in the redox state of peroxiredoxin proteins that is conserved across phylogenetic kingdoms^{4,12}. Critically, this metabolic rhythm persists in the absence of nascent gene expression, both in human cells (anucleate erythrocytes)¹³ and in *Ostreococcus*, which ceases mRNA production upon prolonged photosynthetic inactivity under constant darkness¹⁴, indicating that circadian regulation of cellular metabolism

is not strictly reliant upon rhythmic transcription. These and several other observations render it plausible that circadian rhythms observed in diverse eukaryotes incorporate features of a post-translational timing mechanism that was present in the last eukaryotic common ancestor, LECA¹⁵.

A rich diversity of evolutionarily conserved membrane transporters and channels mediate uptake of ions and micronutrients essential for cellular biochemistry, with several studies having reported their circadian regulation in a variety of contexts (for example ref. 16), and with membrane models of the circadian clock actually predating the identification of any clock genes^{17,18}. It is plausible that rhythmic regulation of ion transport may have conferred an adaptive advantage upon early eukaryotes, allowing the global regulation of biochemical equilibria and reaction rates to tune cellular metabolism with environmental cycles. We therefore asked whether circadian regulation of transmembrane ion transport might constitute a fundamental feature of circadian timekeeping in eukaryotic cells. To find evidence for such regulation in modern organisms, we compared two eukaryotic lineages separated by ~1.5 billion years of evolution⁶: human U2OS cells and *O. tauri*.

Inductively coupled plasma mass spectrometry (ICP-MS) was employed to generate an unbiased analysis of the total cellular elemental composition, including all organelles and cellular structures^{19,20}, over circadian time series. In *Ostreococcus*, clear daily rhythms were detected under natural light/dark cycles for many different ions, including potassium and magnesium (Fig. 1a and Extended Data Fig. 1). Analyses of cells maintained under constant light revealed that, while rhythms of some metal ions ceased under these conditions, oscillations in potassium and magnesium persisted, indicating their regulation by cell-autonomous circadian clock mechanisms. Strikingly, circadian rhythms of magnesium and potassium were also observed in non-proliferating human U2OS cells maintained over three days under constant conditions (Fig. 1b and Extended Data Fig. 2). The oscillation in intracellular potassium is likely to be a consequence of rhythmic Na^+ -dependent pump activity, as circadian regulation of sodium-dependent solute-transport and plasma membrane ATPases has been reported widely (for example refs 2, 16, 21, 22). Calcium is largely found in, and released from, intracellular stores and although calcium has a clearly established role in circadian rhythms¹¹, we infer the absence of any obvious cell-autonomous Ca^{2+} oscillations to mean that its net cellular flux does not vary over the circadian cycle in these cells. We considered the oscillation in Mg^{2+} to be of particular interest, since Mg^{2+} is an essential cofactor for (deoxy-) nucleotide triphosphates. Cellular Mg^{2+} therefore has the potential to regulate many intracellular metabolic reactions, through its requirement for the activities of more than 600 enzymes⁷, including those involved in ATP production, as well as DNA, RNA and protein synthesis. Furthermore, free intracellular Mg^{2+} can act as a second messenger. For example, epidermal growth

¹MRC Laboratory for Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. ²School of Biological Sciences, University of Edinburgh, Max Born Crescent, Edinburgh EH9 3BF, UK. ³Millennium Nucleus for Fungal Integrative and Synthetic Biology, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile. ⁴Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK. ⁵School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, UK.

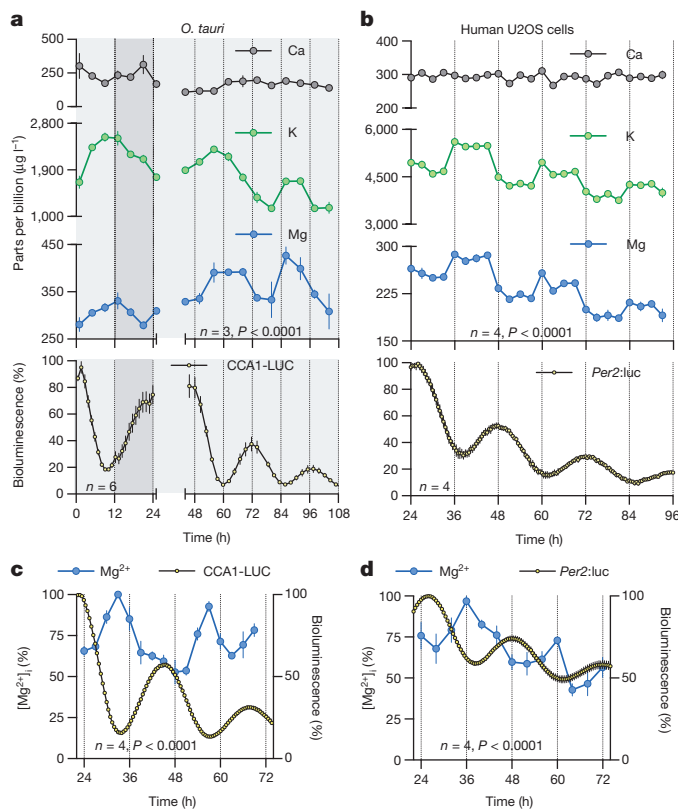


Figure 1 | Conserved cellular rhythms in intracellular magnesium concentrations. **a, b,** Time series of lysates prepared from *Ostreococcus* (**a**, light/dark into constant light) or human U2OS cells (**b**, constant conditions) were analysed by ICP-MS. **c, d,** Rhythms in magnesium concentration in cell lysates were confirmed with luciferase-based assays. Bioluminescence reporters for morning-phased clock gene expression were analysed in parallel (CCA1-LUC and *Per2:luc*) during both assays. All plots show mean \pm s.e.m., with biological replicate numbers (n) indicated. P values report significance by two-way analysis of variance (ANOVA) for time versus element interaction for each pair of elements (**a, b**) or one-way ANOVA (**c, d**).

factor stimulation induces a rapid increase in $[Mg^{2+}]_i$, which acts via highly Mg-sensitive mTOR to activate protein synthesis without any change in total ATP levels²³.

First, we exploited the role of Mg^{2+} as an ATP cofactor, to measure freely available intracellular magnesium concentrations in cellular extracts using the MgATP-dependent enzyme firefly luciferase (Fig. 1c, d). Our initial ICP-MS observations were confirmed using this assay, and we observed clear rhythms of $[Mg^{2+}]_i$ over 2 days under constant conditions in both cell types. Bioluminescent reporters for clock gene activity, recorded in parallel, confirmed the $[Mg^{2+}]_i$ oscillation to occur roughly in antiphase with circadian markers normally expressed around (subjective) dawn (CCA1-LUC, *Per2:luc*, Fig. 1a, b). Conservation between human and algal cells indicates that rhythmic cation transport might constitute a general feature of cellular rhythmicity. We therefore investigated whether such oscillations were also present in the fungus *Neurospora crassa*, representing the third eukaryotic kingdom. Similar to our observations in algal and human cells, a circadian rhythm in cellular magnesium content was observed in antiphase with the abundance of the clock protein FRQ⁹ (Extended Data Fig. 3a–c). We also observed cellular magnesium rhythms in cultured mouse fibroblasts isolated from adult lung, indicating that magnesium rhythms are also present in non-transformed, terminally differentiated mammalian cells (Extended Data Fig. 3d). These rhythms were disrupted in fibroblasts isolated from cryptochrome1^{-/-}, cryptochrome2^{-/-} mice, suggesting their dependence upon clock gene activity. Incidentally, we note that approximately 24 h $[Mg^{2+}]_i$ rhythms

occur cell-autonomously, are temperature-compensated (Extended Data Fig. 4) and entrain to relevant external cues and are therefore circadian by definition²⁴.

The rhythm in total cellular Mg^{2+} measured by ICP-MS must result from daily cycles between net cellular Mg^{2+} influx and efflux, through circadian regulation of plasma membrane Mg^{2+} channel and transporter activity⁷. All known magnesium-transporting proteins in animals (channels TRPM7, MAGT1, MGMT1 and CNMM3, as well as Mg^{2+} -transporter SLC41) exhibit circadian rhythms at the mRNA level in four or more tissues²⁵ (Extended Data Fig. 5). *Ostreococcus* encodes homologues of TRPM7, CNMM3 and SLC41, which are also differentially expressed over the daily cycle (Extended Data Fig. 5). Moreover, short interfering RNA (siRNA)-mediated knockdown of each Mg^{2+} -channel/transporter in U2OS cells results in lengthened circadian period²⁶, suggesting that as well as being clock-regulated, $[Mg^{2+}]_i$ might also feed back to regulate the cellular clock.

To determine whether $[Mg^{2+}]_i$ oscillations are relevant to timekeeping mechanism therefore, we next employed inhibitors of magnesium transport. Cobalt(III)hexammine ($Co(NH_3)_6^{2+}$, CHA) and cobalt(III)chloro-pentammine ($Co(NH_3)_5Cl^{2+}$, CPA) closely resemble a single-solvation shell hydrated Mg^{2+} ion, and have been shown to block Mg^{2+} transport through at least two different transporters/channels^{27,28}. We found both compounds to dose-dependently increase $[Mg^{2+}]_i$ in both cell types (Fig. 2a, b and Extended Data Fig. 6), indicating that these compounds do act to block Mg^{2+} transport. Increased $[Mg^{2+}]_i$ was associated with clear dose-dependent lengthening of circadian period (Fig. 2c–f). Importantly, the effects of CHA on circadian period were dependent on the concentration of extracellular magnesium (Extended Data Fig. 7a–d), indicating a specific role for magnesium in determining the speed at which both algal and human cellular clocks run. To further substantiate this observation, we used quinidine, an inhibitor that acts on several ion transport activities including the SLC41 Na^+/Mg^{2+} antiporter²⁹. Similarly to CHA and CPA, quinidine led to dose-dependent accumulation of intracellular Mg^{2+} and lengthening of circadian period in both cell types (Fig. 2). SLC41 constitutes the sole protein known to exhibit sodium-dependent Mg^{2+} -transport activity²⁹ that is conserved between human and *Ostreococcus* cells. So, to test its specific cellular clock function, we performed siRNA-mediated SLC41 knock-down: observing a clear Mg^{2+} -dependent lengthening of circadian period (Extended Data Fig. 7e–g).

We also observed that depletion of magnesium from the growth medium led to reduced $[Mg^{2+}]_i$, and had dramatic effects on the amplitude and period of the circadian clock in *Ostreococcus* (Fig. 3a, b). Although prolonged growth in low Mg^{2+} media had adverse effects on cell viability of the U2OS line, cells that were simply transferred to Mg^{2+} -free media showed reduced $[Mg^{2+}]_i$, and exhibited circadian rhythms with increased period and decreased bioluminescence amplitude relative to normal media controls (Fig. 3c, d). In neither case was the effect of $[Mg^{2+}]_i$ -depletion attributable to decreased ATP availability, since in both cases cellular ATP levels were significantly increased (Fig. 3b, d).

Thus, as observed previously for cAMP signalling³⁰, treatments that constitutively elevate or reduce $[Mg^{2+}]_i$ both result in period lengthening of clock gene expression in these cell types, indicating that dynamic circadian regulation of $[Mg^{2+}]_i$ might be a cellular clock component. On the other hand, however, it remained possible that Mg^{2+} transport might not contribute to clock mechanism, but instead simply be permissive for cellular timekeeping, analogous to the function of 'housekeeping' genes. To distinguish between these two possibilities, we determined whether an enforced transition in $[Mg^{2+}]_i$ acts as a state variable for cellular circadian oscillations. Upon introduction of magnesium to *Ostreococcus* cells starved of magnesium, we observed strict resetting of the subsequent rhythm to subjective dawn, regardless of prior circadian phase, indicating that changes in $[Mg^{2+}]_i$ can act as a strong zeitgeber (Extended Data Fig. 8). Therefore, our data indicate

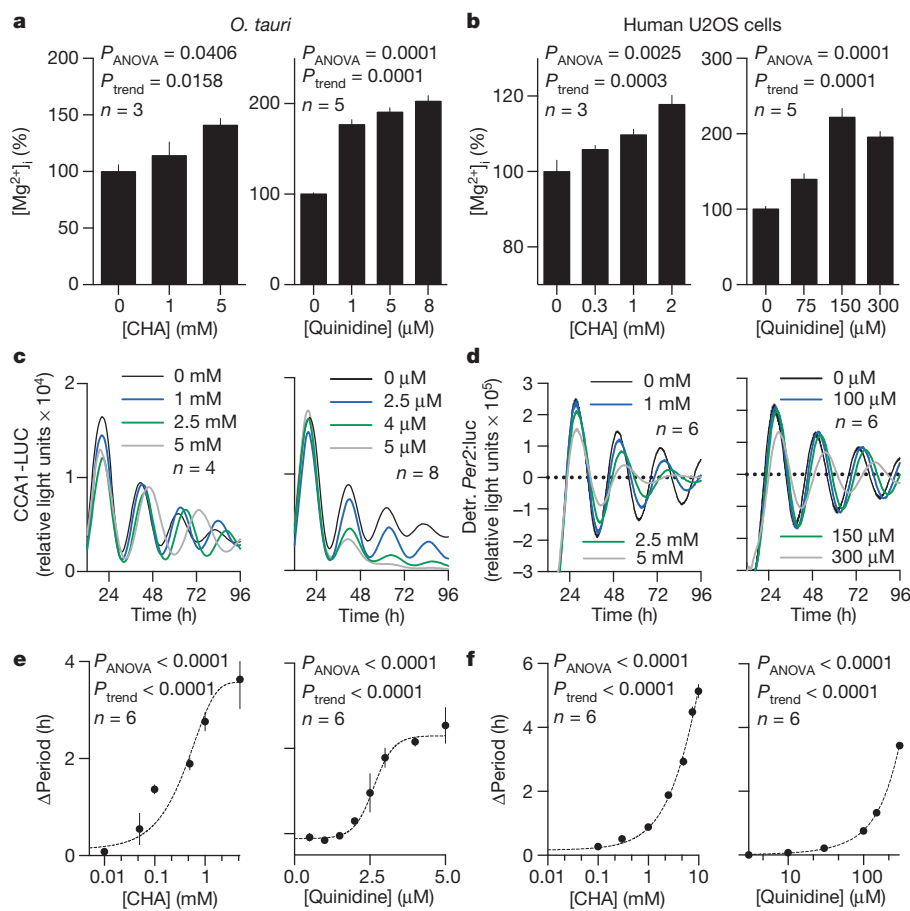


Figure 2 | Chronic inhibition of magnesium transport leads to increased $[Mg^{2+}]_i$ and long circadian period. a–f, Chronic inhibition of magnesium transport by CHA or quinidine increases $[Mg^{2+}]_i$ (a, b) and increases circadian period (c–f). Traces and period dose–response of the CCA1-LUC (*Ostreococcus*) or Per2:luc (U2OS cells) reporters are shown. All plots are mean \pm s.e.m., with biological replicate numbers (n) indicated. P values report significance by one-way ANOVA and post-test for linear trend.

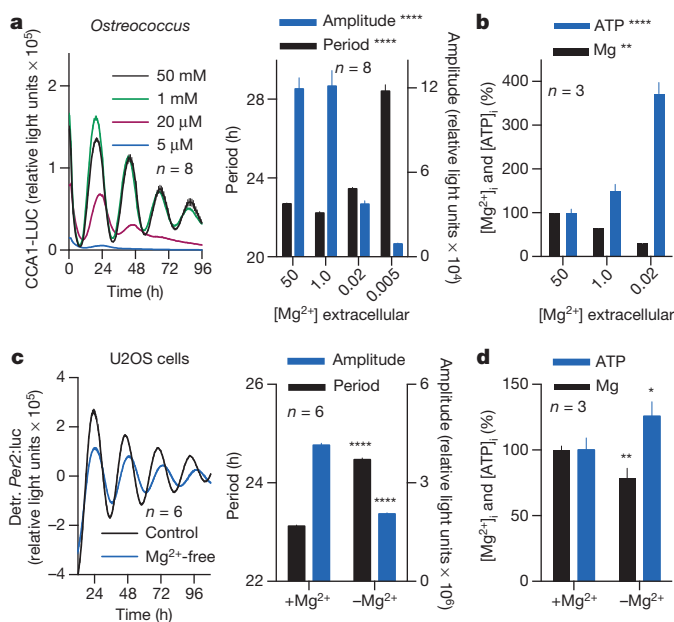


Figure 3 | Reduced $[Mg^{2+}]_i$ affects properties of cellular timekeeping and leads to an increase in ATP. a, c, Bioluminescence traces showing reduced extracellular magnesium significantly affects amplitude and period length of circadian reporters in algal (a) and human cells (c). b, d, Low extracellular magnesium leads to decreased $[Mg^{2+}]_i$ and increased $[ATP]_i$ in both cell types, measured after 4 days. All plots show mean \pm s.e.m., with biological replicate numbers (n) indicated. P values (**** $P < 0.0001$, ** $P = 0.01$, * $P = 0.04$) report significance by one-way ANOVA (*Ostreococcus*) or t -test (U2OS).

that not only does $[Mg^{2+}]_i$ exhibit a bona fide circadian rhythm across diverse eukaryotic cells, but also that appropriate manipulation of $[Mg^{2+}]_i$ is sufficient to determine the key properties of the oscillation (period, amplitude and phase), making $[Mg^{2+}]_i$ indistinguishable from a core clock component.

We considered that the increased cellular ATP levels we observed during Mg^{2+} depletion might be attributable to differential sensitivity of MgATP-dependent cytosolic enzymes compared with the organellar ATP synthesis machinery. For example, ATP accumulation was accompanied by a marked reduction in extracellular lactate accumulation in U2OS cultures (Extended Data Fig. 9a), indicative of reduced glycolysis. We therefore considered whether rate-changes in gross cytosolic energy metabolism might be a functional consequence of cell-autonomous circadian $[Mg^{2+}]_i$ oscillations. A clear prediction would be that global rates of translation should be limited at circadian phases of low $[Mg^{2+}]_i$, since protein synthesis is one of the most energetically expensive processes that cells undertake.

We assayed translation rate by puromycin incorporation⁸ in both cell types just before (anticipated) biological dusk and dawn; at the phase of lowest and highest $[Mg^{2+}]_i$, respectively. The *Ostreococcus* experiment was performed under its natural light/dark cycle so as to best model an organism in its natural environment, whereas the U2OS experiment was performed under constant conditions to model innate peripheral cellular clock function. We observed that both *Ostreococcus* and U2OS cells did exhibit significantly higher translation rates at the peak of $[Mg^{2+}]_i$, as predicted (Fig. 4c, d). In mammalian cells, the highly MgATP-sensitive mTOR pathway was recently shown to mediate circadian control of translation⁸. We hypothesized that $[Mg^{2+}]_i$ oscillations might act via mTOR to effect circadian translational regulation, and tested this using two pharmacologically distinct mTOR inhibitors (torin1 and rapamycin). Both inhibitors lengthened period dose-dependently, and abolished any additional period lengthening due to depletion of extracellular magnesium that was

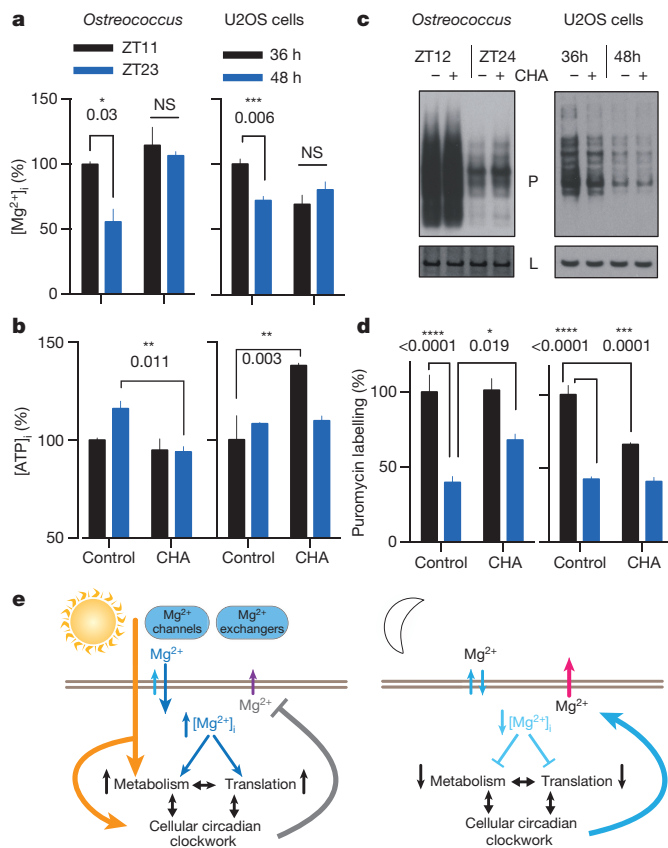


Figure 4 | Magnesium transport phase-dependently affects cellular energy balance. **a–d**, Cells were treated with CHA or vehicle during the second half of the day (black bars) or night (blue bars) to test acute effects on magnesium (**a**) and ATP (**b**) levels in both cell types. Incorporation of puromycin was analysed by western blot (**c**) then quantified to give relative translation rates at the indicated times (**d**). Mean \pm s.e.m. are shown, $n = 3$ biological replicates, t -test significance reported (NS, $P > 0.18$). **e**, Simplified model for a feedback mechanism between circadian [Mg²⁺]_i rhythms and the clockwork to temporally regulate global cellular metabolism; orange arrow represents extracellular stimuli.

observed in controls. This clear ‘ceiling effect’ strongly suggests changes in [Mg²⁺]_i act through mTOR activity to regulate cellular circadian period (Extended Data Fig. 9b–e). To test the extent to which marked differences in translation rates between dawn and dusk were attributable to cell-autonomous [Mg²⁺]_i oscillations, we incubated cells acutely with CHA to block magnesium transport before assaying overall translational rates. CHA significantly attenuated the difference in [Mg²⁺]_i between dawn and dusk (Fig. 4a), and phase-dependently affected ATP levels (Fig. 4b). Crucially, differential translation rates were attenuated similarly by CHA treatment (Fig. 4c, d), indicating that in both cell types, circadian regulation of magnesium levels contributes to circadian rhythms in global translation rate.

Our data support a model (Fig. 4e and Extended Data Fig. 10a) where the cellular clockwork regulates the expression of plasma membrane Mg²⁺ channels and transporters to generate rhythmic magnesium fluxes. These rhythms appear to facilitate the higher energetic demands and protein production of human cells during the (biological) day, as well as the global downregulation of ATP turnover and translation in photosynthetic cells at night. Reciprocally, the [Mg²⁺]_i rhythm feeds back to regulate the period, phase and amplitude of clock gene expression rhythms, acting as a ‘meta-regulator’ to integrate metabolic rhythms with transcriptional feedback models of cellular timekeeping. It is noteworthy, however, that the magnesium rhythm observed in *Ostreococcus* persists under transcriptionally inactive conditions¹⁴ of constant darkness (Extended Data Fig. 10b, c), in the absence of

transcriptional regulation of membrane transport. It is therefore likely that some aspects of circadian magnesium flux are regulated by, and may contribute to, the same uncharacterized non-transcriptional clock mechanism^{13,14} that also drives persistent peroxiredoxin overoxidation rhythms in transcriptionally inactive cells across taxa, and which we speculate was present in the LECA.

Cell-autonomous rhythms in [Mg²⁺]_i availability have the potential to impart circadian regulation to any cellular system where MgNTP hydrolysis becomes rate limiting. Although the clinical relevance of [Mg²⁺]_i in various tissues is beginning to garner more attention, the interactions between magnesium transport and human health are poorly understood. Further investigation of the downstream consequences of circadian regulation of [Mg²⁺]_i will therefore be important.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 December 2015; accepted 8 February 2016.

Published online 13 April 2016.

- Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A. & Harmer, S. L. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol.* **9**, R130 (2008).
- Hughes, M. E. *et al.* Harmonics of circadian gene transcription in mammals. *PLoS Genet.* **5**, e1000442 (2009).
- Endo, M., Shimizu, H., Nohales, M. A., Araki, T. & Kay, S. A. Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature* **515**, 419–422 (2014).
- Edgar, R. S. *et al.* Peroxiredoxins are conserved markers of circadian rhythms. *Nature* **485**, 459–464 (2012).
- Bass, J. Circadian topology of metabolism. *Nature* **491**, 348–356 (2012).
- Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
- de Baaij, J. H., Hoenderop, J. G. & Bindels, R. J. Magnesium in man: implications for health and disease. *Physiol. Rev.* **95**, 1–46 (2015).
- Lipton, J. O. *et al.* The circadian protein BMAL1 regulates translation in response to S6K1-mediated phosphorylation. *Cell* **161**, 1138–1151 (2015).
- Dunlap, J. C. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
- Corellou, F. *et al.* Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote *Ostreococcus*. *Plant Cell* **21**, 3436–3449 (2009).
- Hastings, M. H., Maywood, E. S. & O'Neill, J. S. Cellular circadian pacemaking and the role of cytosolic rhythms. *Curr. Biol.* **18**, R805–R815 (2008).
- Olmedo, M. *et al.* Circadian regulation of olfaction and an evolutionarily conserved, nontranscriptional marker in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **109**, 20479–20484 (2012).
- O'Neill, J. S. & Reddy, A. B. Circadian clocks in human red blood cells. *Nature* **469**, 498–503 (2011).
- O'Neill, J. S. *et al.* Circadian rhythms persist without transcription in a eukaryote. *Nature* **469**, 554–558 (2011).
- van Ooijen, G. & Millar, A. J. Non-transcriptional oscillators in circadian timekeeping. *Trends Biochem. Sci.* **37**, 484–492 (2012).
- Ko, G. Y., Shi, L. & Ko, M. L. Circadian regulation of ion channels and their functions. *J. Neurochem.* **110**, 1150–1169 (2009).
- Njus, D., Sulzman, F. M. & Hastings, J. W. Membrane model for the circadian clock. *Nature* **248**, 116–120 (1974).
- Nitabach, M. N., Holmes, T. C. & Blau, J. Membranes, ions, and clocks: testing the Njus-Sulzman-Hastings model of the circadian oscillator. *Methods Enzymol.* **393**, 682–693 (2005).
- Danku, J. M., Lahner, B., Yakubova, E. & Salt, D. E. Large-scale plant ionomics. *Methods Mol. Biol.* **953**, 255–276 (2013).
- Danku, J. M. C., Gumaelius, L., Baxter, I. & Salt, D. E. A high-throughput method for *Saccharomyces cerevisiae* (yeast) ionomics. *J. Anal. At. Spectrom.* **24**, 103–107 (2009).
- Nishinaga, H. *et al.* Circadian expression of the Na⁺/H⁺ exchanger NHE3 in the mouse renal medulla. *Biomed. Res.* **30**, 87–93 (2009).
- Wang, Y. C., Chen, Y. S., Cheng, R. C. & Huang, R. C. Role of Na⁺/Ca²⁺ exchanger in Ca²⁺ homeostasis in rat suprachiasmatic nucleus neurons. *J. Neurophysiol.* **113**, 2114–2126 (2015).
- Rubin, H. The logic of the membrane, magnesium, mitosis (MMM) model for the regulation of animal cell proliferation. *Arch. Biochem. Biophys.* **458**, 16–23 (2007).
- Pittendrigh, C. S. Circadian rhythms and the circadian organization of living systems. *Cold Spring Harb. Symp. Quant. Biol.* **25**, 159–184 (1960).
- Pizarro, A., Hayer, K., Lahens, N. F. & Hogenesch, J. B. CircaDB: a database of mammalian circadian gene expression profiles. *Nucleic Acids Res.* **41**, D1009–D1013 (2013).
- Zhang, E. E. *et al.* A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* **139**, 199–210 (2009).

27. Kolisek, M. *et al.* SLC41A1 is a novel mammalian Mg^{2+} carrier. *J. Biol. Chem.* **283**, 16235–16247 (2008).
28. Kucharski, L. M., Lubbe, W. J. & Maguire, M. E. Cation hexaamines are selective and potent inhibitors of the CorA magnesium transport system. *J. Biol. Chem.* **275**, 16767–16773 (2000).
29. Kolisek, M., Nestler, A., Vormann, J. & Schweigel-Rontgen, M. Human gene SLC41A1 encodes for the Na^+/Mg^{2+} exchanger. *Am. J. Physiol. Cell Physiol.* **302**, C318–C326 (2012).
30. O'Neill, J. S., Maywood, E. S., Chesham, J. E., Takahashi, J. S. & Hastings, M. H. cAMP-dependent signaling as a core component of the mammalian circadian pacemaker. *Science* **320**, 949–953 (2008).

Acknowledgements G.v.O. is supported by a Royal Society University Research Fellowship (UF110173) and research grants (RS120372 and RS140275). J.S.O. is supported by the Medical Research Council (MC_UP_1201/4) and the Wellcome Trust (093734/Z/10/Z). M.P. is funded by KWF BUIT 2014-6637. L.F.L. and C.O.Y. are supported by Millennium Nucleus for Fungal Integrative and Synthetic Biology (NC120043), and Fondo Nacional de Desarrollo Científico y

Tecnológico (FONDECYT 1131030). S. K. Hodge is acknowledged for managing imaging facilities for algal experiments. At the MRC LMB, we are grateful to the Biomedical Services Group for animal care; M. Hastings and J. Chesham for supplying mouse tissue; P. Margiotta for assistance with figures. We also thank D. E. Salt, M. Knight, E. Grünwald, P. Crosby, L. Hewitt and B. Cross for criticism. The anti-puromycin ascites was a gift from M. Hegde.

Author Contributions G.v.O. and J.S.O. conceived the approach and designed the study. L.F.L. and C.O.Y. generated the *Neurospora* result. J.D. and L.E. performed ICP analyses. G.v.O. and L.L.H. performed *Ostreococcus* experiments. Human U2OS cell experiments were performed by K.A.F. M.P. performed mouse fibroblast experiments. N.P.H. provided analytical and intellectual contributions. G.v.O. and J.S.O. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.v.O. (Gerben.vanOoijen@ed.ac.uk) or J.S.O. (oneillj@mrc-lmb.cam.ac.uk).

METHODS

The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

All materials were purchased from Sigma-Aldrich unless otherwise stated. ICP-MS data are reported as 'parts per billion ($\mu\text{g l}^{-1}$)', Mg^{2+} measured by luciferase assay is normalized to the highest value and reported as ' $[\text{Mg}^{2+}]_i$ '.

***Ostreococcus tauri*.** Wild-type cells or cells transgenically expressing a translational fusion of CCA1 to luciferase from the CCA1 promoter (CCA1-LUC)¹⁰ were grown under 12 h/12 h light/dark cycles in artificial sea water (24 g l^{-1} NaCl, 4 g l^{-1} Na_2SO_4 , 0.68 g l^{-1} KCl, 200 mg l^{-1} NaHCO_3 , 100 mg l^{-1} KBr, 25 mg l^{-1} H_3BO_3 , 3 mg l^{-1} NaF, plus hydrous salts: $50 \text{ mM MgCl}_2 \cdot 6\text{H}_2\text{O}$, $10 \text{ mM CaCl}_2 \cdot 2\text{H}_2\text{O}$, $0.1 \mu\text{M SrCl}_2 \cdot 6\text{H}_2\text{O}$), supplemented with Guillard's F/2 marine enrichment solution and $10 \text{ nM H}_2\text{SeO}_3$. Full medium was adjusted to a salinity of 30 p.p.t.

Imaging and analysis of luminescent rhythms were performed as described^{31–33}. For resetting experiments, magnesium-free media were removed with a multichannel pipette and replaced with magnesium-containing media. In all luminescent imaging experiments, eight replicate wells constitute $n = 8$, and presented experiments are representative of three or more replicate experiments.

For ICP-MS analyses, 30 ml culture was pelleted, washed three times in 1 M sorbitol to remove sea water, and digested in $100 \mu\text{l}$ nitric acid (69%, ARISTAR grade, VWR International) spiked with $345 \text{ p.p.b. indium}$ (VWR International) at room temperature ($\sim 20^\circ\text{C}$) for $\sim 3 \text{ h}$. Samples were then diluted to a final concentration of 2% v/v nitric acid and 10 p.p.b. indium before analysis on an Agilent 7500ce with octopole reaction system. Serial dilutions of ICP-Multi-element solution IV (Merck, Certipur) were used for calibration of all the metals analysed and to check for instrument drift. A standard reference material, SRM1643e (National Institute of Standards and Technology), was analysed to validate the calibration. Indium was used to correct for dilution errors introduced during handling. ICP-MS data reported are based on three replicate flasks, each sampled every time point ($n = 3$). Results presented were verified in a replicate experiment, and outliers were excluded if they were $> 2 \text{ s.d.}$ from the mean.

Cell extracts for luminescent Mg^{2+} and ATP assays were made from three replicates ($n = 3$) of 5 ml cell culture, pelleted and washed with 1 M sorbitol, and resuspended in $100 \mu\text{l}$ medium before adding $100 \mu\text{l}$ $2\times$ extraction buffer (1% Triton X-100, 300 mM NaCl, 100 mM HEPES). Twenty-five microlitres of extract was boiled and added to $75 \mu\text{l}$ of assay mix (40 mM HEPES, 1 mM luciferin, 0.05 mg ml^{-1} QuantiLum (Promega), and either 1 mM MgCl_2 or $10 \mu\text{M}$ ATP). Luminescence was measured on a TopCount (Packard) plate reader against a standard curve. As Mg^{2+} ions that remain tightly bound to cellular macromolecules such as membrane components and DNA are not detected by this alternative assay, the relative amplitude of $[\text{Mg}^{2+}]_i$ changes observed using this assay were substantially larger than measured by ICP-MS. Quinidine, CHA and CPA were made up in medium and added 24 h before cell lysis for chronic treatments. Results were verified in one or more replicate experiments. For puromycin experiments (Fig. 4), cobalt amines or vehicle were added at ZT6 or ZT18, and 0.5 mg ml^{-1} puromycin was added 20 min before harvesting cells at ZT11 and ZT23 (required concentration and incubation time determined empirically to reduce expression from a constitutive promoter driving luciferase by about half). Analysis of incorporation was performed as described below for U2OS cells. Loading control was RbL (Coomassie).

To identify potential *Ostreococcus* transporter proteins, mammalian sequences for all classes of SLC and all known magnesium transporters were blasted onto the *Ostreococcus* proteome using DELTA-BLAST (National Center for Biotechnology Information), and gene models were then taken from the latest version of the *Ostreococcus* genome³⁴ using the Orcae service³⁵ (Ghent University).

***Neurospora crassa*.** Plates with 25 ml of Vogel's medium containing 2% glucose, 0.5% arginine, 10 ng ml^{-1} biotin and 0.2% Tween 80 were inoculated with 10^6 conidia of wild-type *ras-1^{bd}* strain and incubated under constant light for 48 h at 30°C . Two 2-cm disks cut from the mycelial pad were placed in a series of 50 ml cultures (Vogel's medium containing 0.03% glucose, 0.05% arginine and 10 ng ml^{-1} biotin). These cultures were incubated at 25°C under constant light before staggered transfers at 4 h intervals to constant darkness, shaking at 125 r.p.m. Mycelia were then washed four times in a falcon tube with 25 ml of sterile water, dried on filtration paper and frozen in liquid N_2 , and stored at -80°C . Samples were then lyophilized and ground to a powder. These samples were split into two: one for western blot analysis of FRQ oscillations (as described in ref. 36) and one for ICP-MS analysis. ICP-MS was performed as described for *Ostreococcus*, except that the dissolved tissue was filtered through a $0.22 \mu\text{m}$ filter before application on ICP-MS.

Mammalian cells. Human U2OS cells, purchased from ATCC, were stably transfected with *Per2:luc* and cultured between passage numbers 31–51 as described previously³⁷, except that 10% FetalClone II serum (HyClone) was used

in place of fetal bovine serum. Cells were mycoplasma-free (Mycoalert, Lonza) and authenticated by Multiplex PCR. Cells were seeded into 6-, 24- or 96-well white plates at a density of 10^5 cells per millilitre, and incubated in a humidified incubator (37°C , 5% CO_2) under 12 h:12 h 32.37°C temperature cycles until confluent for a minimum of 3 days. We confirmed previous reports that U2OS do not proliferate appreciably once confluent and that, in the presence of B-27 supplement, U2OS circadian period is not affected by the presence of additional serum during bioluminescence recordings (Extended Data Fig. 2b–d). We cannot formally exclude the rhythmic excretion of growth factors. Empirically we determined that, upon transfer to constant 37°C , *Per2:luc* bioluminescence peaks around the anticipated transition from 32 to 37°C , and 28 h after a media/serum change. To maximize intercellular synchrony therefore, for all U2OS experiments, media was changed at 4 h before the warm phase for HEPES-buffered 'air medium'³⁸ and then maintained at constant 37°C under a gas-impermeable seal. Air medium stock was prepared as described previously³⁸ and supplemented with 2% B-27 (Life Technologies, $50\times$), $300 \mu\text{M}$ luciferin (Biosynth AG), 1% glutamax (Life Technologies), 100 units ml^{-1} penicillin/ $100 \mu\text{g ml}^{-1}$ streptomycin, as well as FetalClone II serum (HyClone). Serum was present at 10% for every U2OS experiment except those performed with Mg-free air medium (and controls). Mg-free air medium was prepared from its individual components, replacing magnesium with sodium. The final osmolarity was adjusted to 350 mOsm with NaCl and sterile filtered. Bioluminescence recordings were performed in a lumicycle (Actimetrics), a LB962 plate reader (Berthold technologies) or an Alligator (Cairn Research). All bioluminescence data were independently verified by at least two replicate experiments. Primary fibroblasts were isolated from the lung tissue of adult cryptochrome1^{-/-}, cryptochrome2^{-/-} male mice³⁹ or wild-type controls, and cultured as described previously³⁸, then immortalized by serial passage⁴⁰. All animal work was licensed under the UK Animals (Scientific Procedures) Act 1986 with local ethical approval. For time courses, confluent cultures were instead synchronized by incubation for 2 h with 100 nM dexamethasone, then changed to air medium supplemented as described above, but containing 1% FetalClone III (instead of II) serum (HyClone).

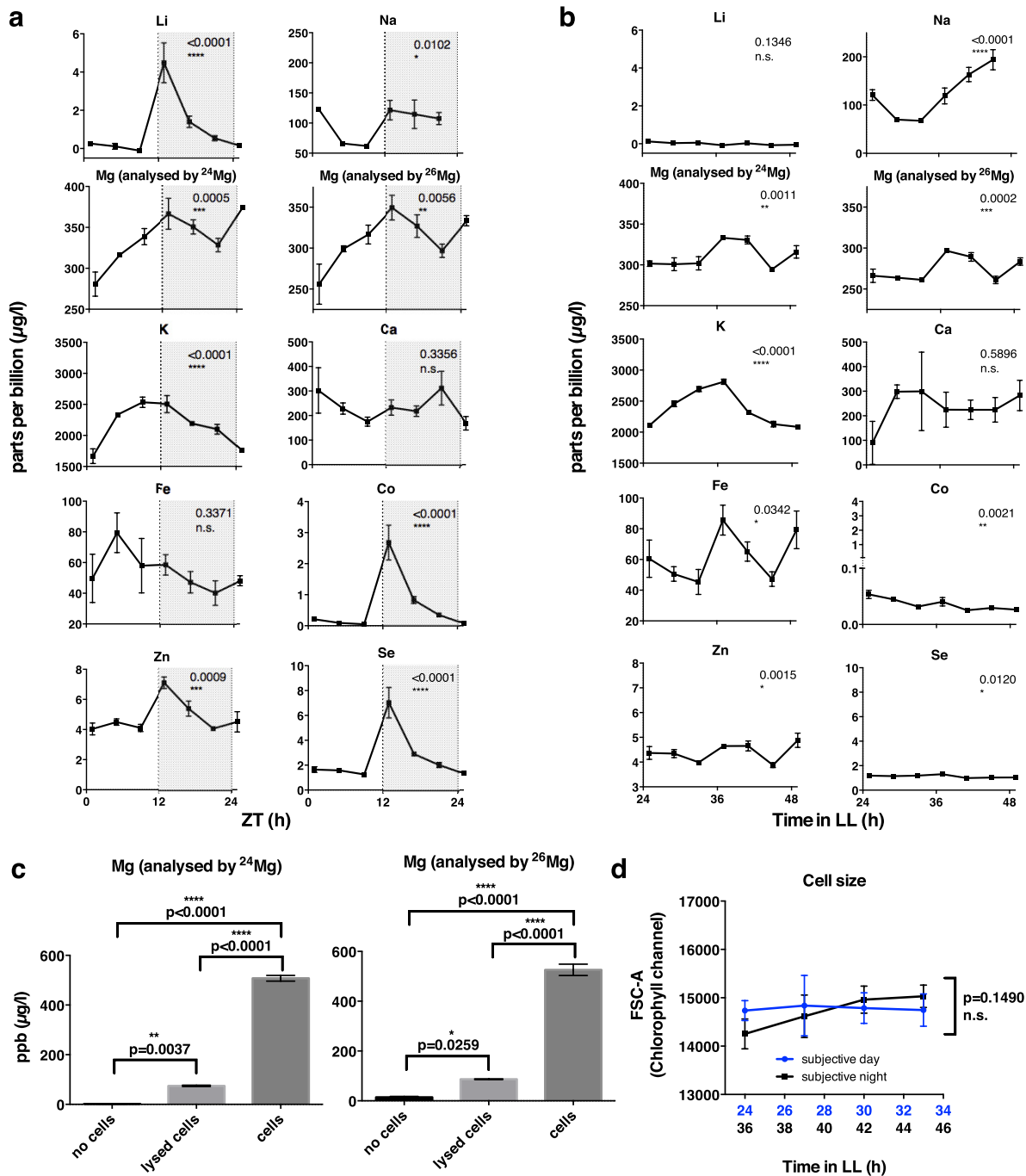
For ICP-MS analyses, cells were washed twice in a salt-free isosmotic buffer (300 mM sucrose, 10 mM Tris pH7.4, 350 mOsm) at room temperature to remove ions in the cell media. Cells were then digested in 65% nitric acid supplemented with 100 p.p.b. cerium for 30 min at room temperature and flash frozen. Upon thaw, samples were heated at 90°C for 1 h then centrifuged at 18,000g for 20 min to remove any debris, then diluted 1:12 in HPLC-grade water to give a final matrix concentration of 5% HNO_3 . The ICP-MS time course data shown are representative of three separate time courses, all with the indicated number of biological replicates per time point. ICP-MS was performed on a Perkin Elmer Elan DRC II. SPS-SW2 (LGC) was used as a routine standard but linear sensitivity was confirmed for each element (Extended data Fig. 2). Cerium in the HNO_3 extraction reagent was used to correct for dilution errors introduced during handling. Outliers were excluded if they were $> 2 \text{ s.d.}$ from the mean.

For intracellular ATP and Mg^{2+} assays, U2OS cells were washed twice with ice-cold PBS + 5 mM EDTA then lysed in buffer containing 30 mM HEPES pH 7.4, 100 mM NaCl, 10% glycerol, 1% Triton X-100 and 5 mM Na_3VO_4 , then flash frozen. Mouse fibroblasts were instead lysed in 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% LDS, 1% Triton X-100, 0.5% Na-deoxycholate and protease inhibitors. Samples were thawed, and proteins denatured, by heating at 90°C for 10 min followed by centrifugation at 18,000g at 4°C . Assays were performed by diluting samples 1:1 in $2\times$ assay buffer giving final concentrations of 30 mM HEPES (pH 7.4), 1 mM luciferin, 50 nM QuantiLum (Promega) and either 15 mM MgSO_4 or $100 \mu\text{M}$ ATP (dependent on whether ATP or Mg^{2+} was being assayed). All results were independently verified in at least two replicate experiments. For Mg^{2+} assays upon mouse fibroblast lysates, 5% bovine serum albumin was also included in the assay buffer to quench LDS. Bioluminescence was measured using a LB962 CentroPRO microplate reader (Berthold Technologies) or a Spark 10M microplate reader (Tecan). Lactate assay kits (MAK064-1KT) were used in accordance with the manufacturer's instructions and measured on a Spark 10M microplate reader (Tecan). SLC41A1 siRNA (Santa Cruz sc-88707) or control siRNA-A (Santa Cruz sc-37007) was used at 80 pM to $6 \mu\text{l}$ of transfection agent (Santa Cruz sc-29528) to transfect *Per2:luc* U2OS cells at 60% confluence in 35 mm^2 dishes according to the manufacturer's instructions. Quinidine, rapamycin, torin1, CHA and CPA stock solutions were made up in serum and B-27-free air medium, with chronic treatments lasting for 24 h before enzymatic $[\text{Mg}^{2+}]_i$ assay. For acute treatments, 2 mM CHA (final concentration) was added 3 h before the peak or trough of *Per2:luc* bioluminescence and harvested 3 h later. Puromycin labelling was performed for 10 min at 37°C with $10 \mu\text{g ml}^{-1}$ puromycin before cells were washed twice in ice-cold PBS + 5 mM EDTA then lysed on ice for 20 min in buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Triton X-100

and 0.5% sodium deoxycholate. Gel electrophoresis, transfer and blotting were performed as described previously⁴. SLC41A1 rabbit antibody (Abcam ab83701) was diluted by 1:1,000, and mouse monoclonal anti-puromycin ascites⁴¹ was diluted 1:100 in blocking buffer (w/v 0.25% BSA, 0.25% dried skimmed milk in Tris-buffered saline/0.05% Tween-20) and incubated overnight with nitrocellulose membranes. Anti- β actin (Santa Cruz, sc-47778) was used as a loading control at 1:5,000 dilution. U2OS impedance measurements were performed on an xCELLigence RTPA DP under our standard bioluminescence recording conditions (air media with supplements, including 10% serum) according to the manufacturer's instructions.

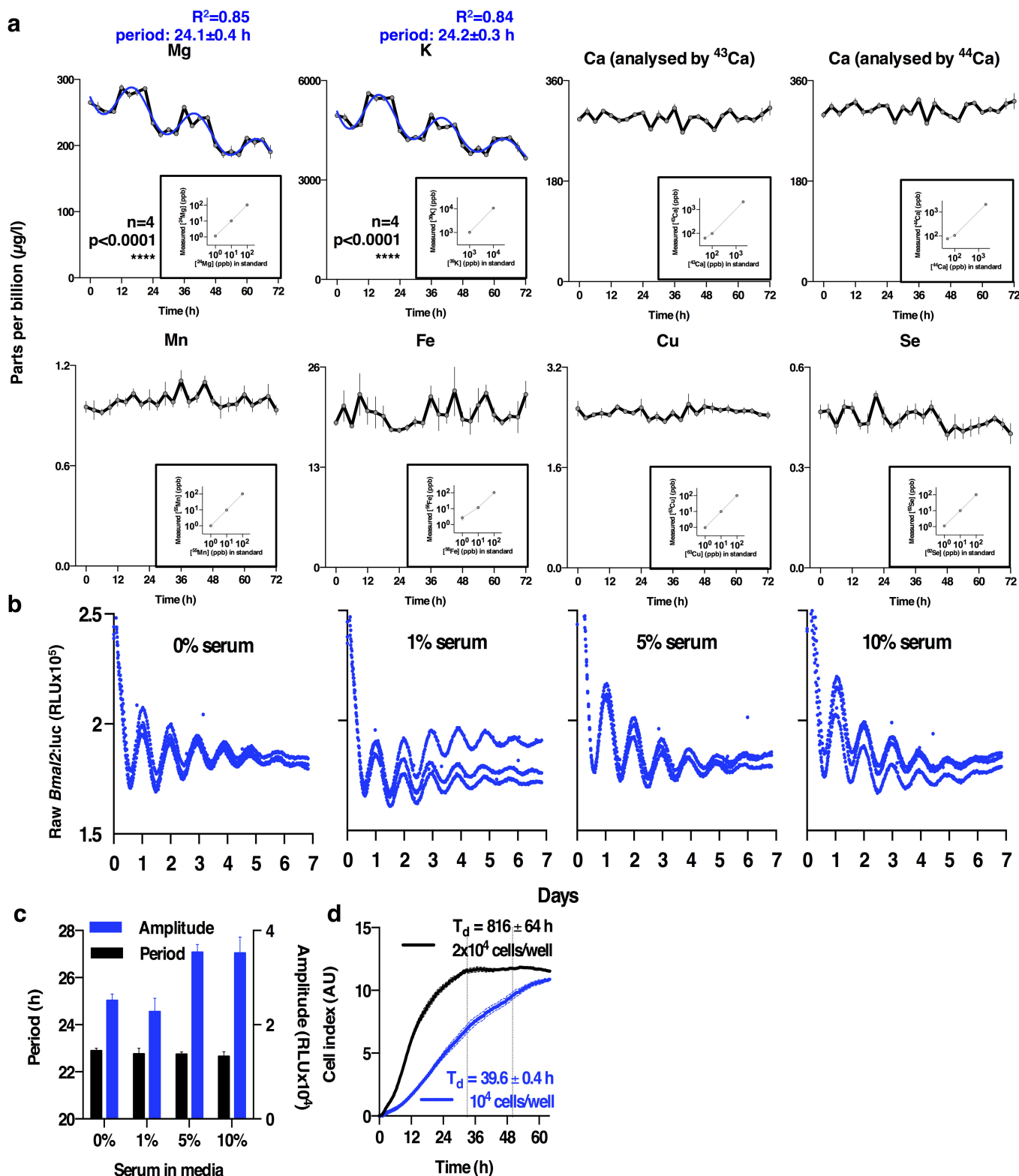
Analysis. Statistical tests were performed using Graphpad Prism, with all tests two-sided. No significant difference in variance was detected between groups under comparison (Brown–Forsythe's test; $P > 0.2$). Separate power calculations were not performed for every experiment. Numbers of biological replicates were instead chosen on the basis of preliminary experiments, so that an effect size of at least 10% could be detected between experimental groups for $\alpha = 0.05$ and $\beta = 0.9$. For phase response curves, the circular mean and s.d. were calculated using the 'circular' R package implementing the methods described previously^{42,43}. A least squared fit of the phase response versus phase before addition to a linear model where the gradient was constrained to -1 was performed. The y intercept where $x = 0$ is indicative of the nascent phase relative to the prior phase in hours.

31. van Ooijen, G., Dixon, L. E., Troein, C. & Millar, A. J. Proteasome function is required for biological timing throughout the twenty-four hour cycle. *Curr. Biol.* **21**, 869–875 (2011).
32. van Ooijen, G. *et al.* Functional analysis of casein kinase 1 in a minimal circadian system. *PLoS ONE* **8**, e70021 (2013).
33. Le Bihan, T. *et al.* Label-free quantitative analysis of the casein kinase 2-responsive phosphoproteome of the marine minimal model species *Ostreococcus tauri*. *Proteomics* **15**, 4135–4144 (2015).
34. Blanc-Mathieu, R. *et al.* An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genom.* **15**, 1103 (2014).
35. Sterck, L., Billiau, K., Abeel, T., Rouze, P. & Van de Peer, Y. ORCAE: online resource for community annotation of eukaryotes. *Nature Methods* **9**, 1041 (2012).
36. Baker, C. L., Kettenbach, A. N., Loros, J. J., Gerber, S. A. & Dunlap, J. C. Quantitative proteomics reveals a dynamic interactome and phase-specific phosphorylation in the *Neurospora* circadian clock. *Mol. Cell* **34**, 354–363 (2009).
37. Valekunja, U. K. *et al.* Histone methyltransferase MLL3 contributes to genome-scale circadian transcription. *Proc. Natl Acad. Sci. USA* **110**, 1554–1559 (2013).
38. O'Neill, J. S. & Hastings, M. H. Increased coherence of circadian rhythms in mature fibroblast cultures. *J. Biol. Rhythms* **23**, 483–488 (2008).
39. van der Horst, G. T. *et al.* Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398**, 627–630 (1999).
40. Xu, J. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) Ch. 28, Unit 28.1 (Wiley, 2005).
41. David, A. *et al.* Nuclear translation visualized by ribosome-bound nascent chain puromycylation. *J. Cell Biol.* **197**, 45–57 (2012).
42. Fisher, N. I. *Statistical Analysis of Circular Data*, xviii + 1–277 (Cambridge Univ. Press, 1993).
43. Jammalamadaka, S. R. & SenGupta, A. *Topics in Circular Statistics* 1–322 (World Scientific, 2001).
44. Hirota, T. *et al.* A chemical biology approach reveals period shortening of the mammalian circadian clock by specific inhibition of GSK-3 β . *Proc. Natl Acad. Sci. USA* **105**, 20746–20751 (2008).
45. Monnier, A. *et al.* Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. *BMC Genom.* **11**, 192 (2010).
46. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10**, R130 (2009).
47. Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct* **7**, 12 (2012).



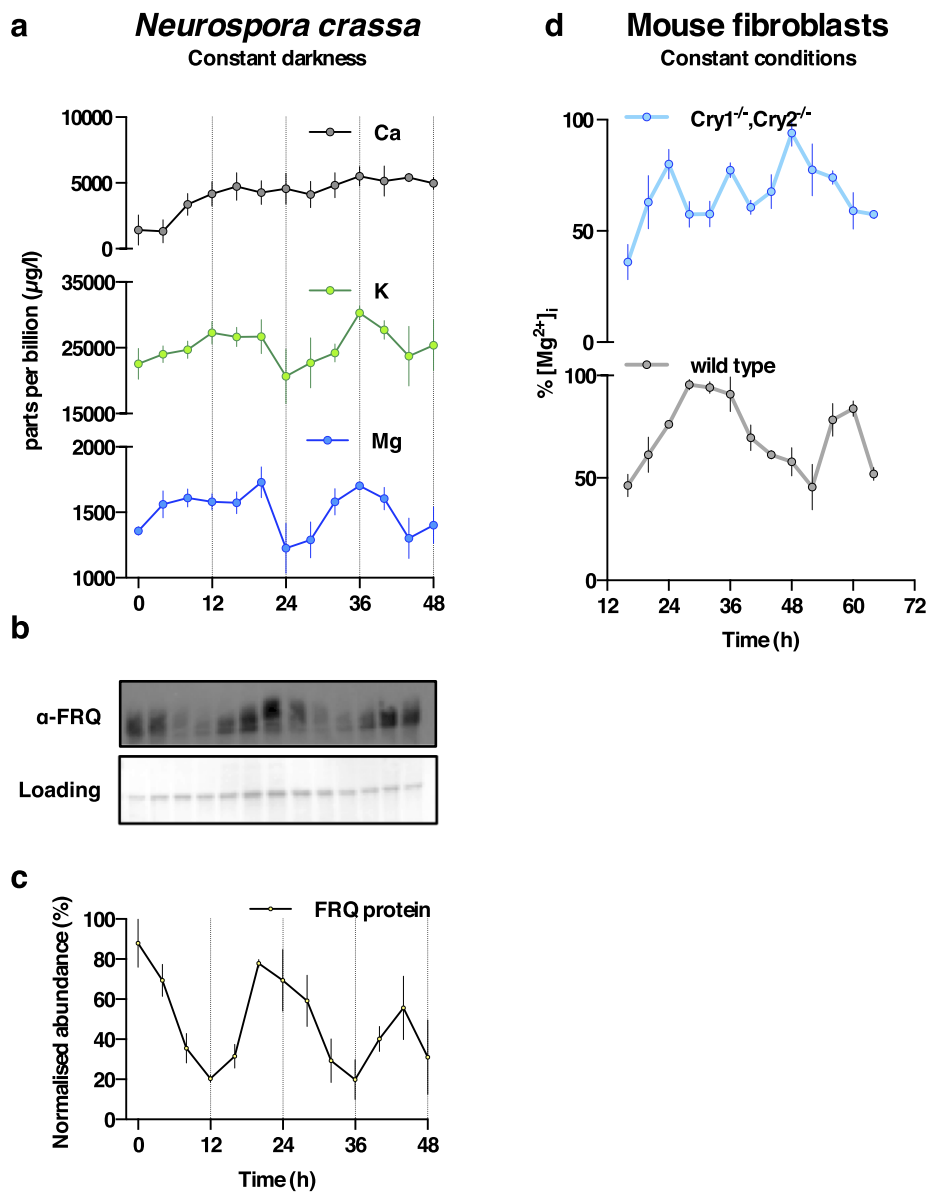
Extended Data Figure 1 | Additional ICP-MS data and controls (*Ostreococcus*). **a, b**, ICP-MS analyses of cell lysates from 12 h:12 h light/dark cycles (**a**) or on the second day of constant light (**b**). *P* values report significance by one-way ANOVA (mean \pm s.e.m., $n=3$). **c**, ICP-MS analyses on cell lysates compared with media control (no cells) and membrane fractions (lysed cells) (mean \pm s.d. plotted, $n=2$ biological replicates), indicating that magnesium signal in **a** and Fig. 1 comes

predominantly from the intracellular space. Groups are significantly different by one-way ANOVA ($P < 0.0001$); Tukey's multiple comparison *P* values are indicated. **d**, Fluctuations in measured concentrations are not related to fluctuations in cell size over time. No significance of time as source of variation in cell size was observed by fluorescence-activated cell sorting analyses (mean \pm s.d. plotted, one-way ANOVA *P* value is indicated, $n=5$ biological replicates).



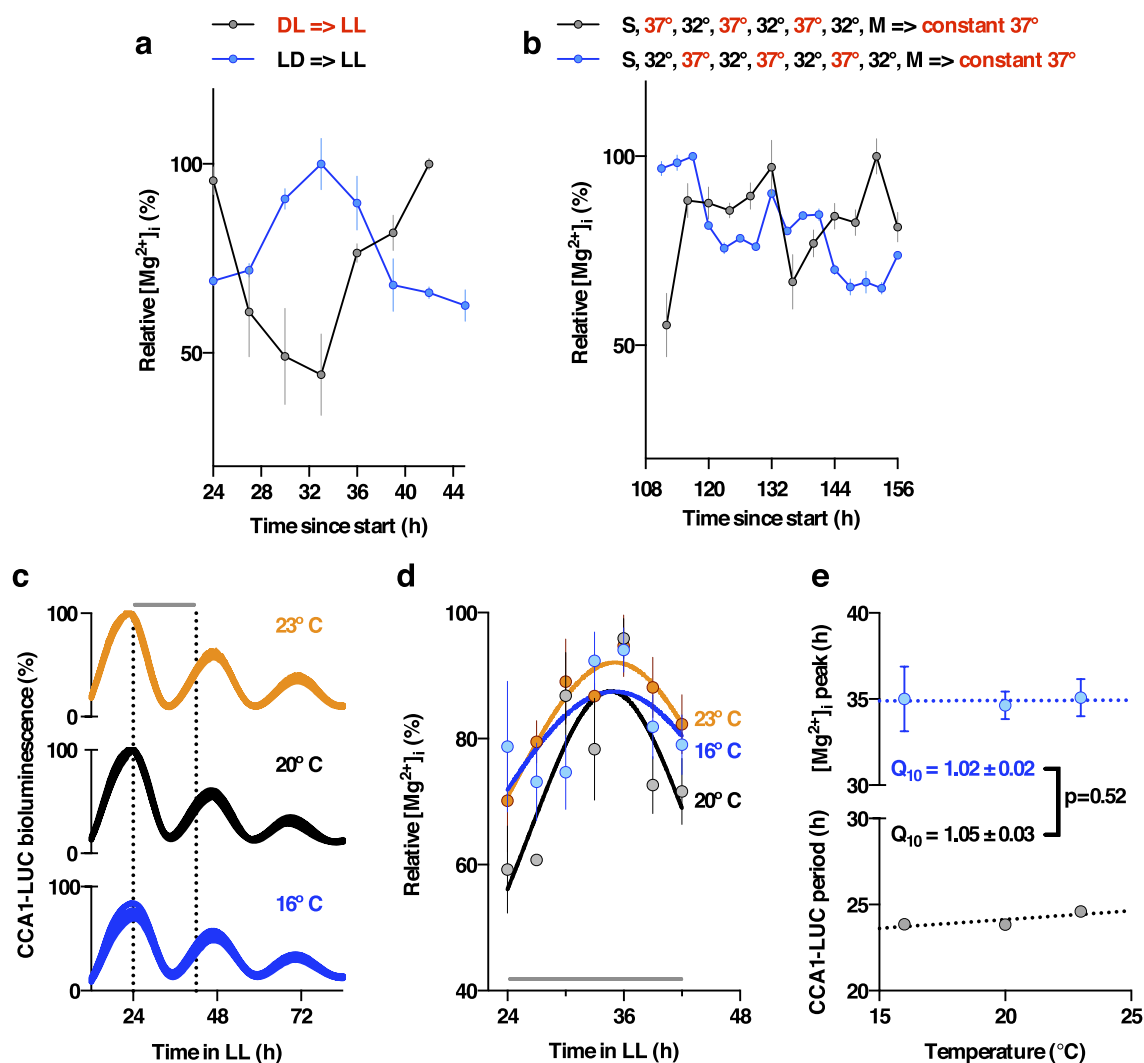
Extended Data Figure 2 | Additional ICP-MS data (U2OS cells) and controls. **a**, ICP-MS analyses of U2OS cell extracts for stable isotopes of several biologically relevant ions (mean \pm s.e.m., grey/black, $n \geq 4$ biological replicates), with insets showing standards that indicate linearity over the observed concentration ranges (mean \pm percentage coefficient of variation). We compared how well a straight-line + damped sine wave model (adapted from ref. 44) fit to each time series compared with a straight-line only (null hypothesis, no rhythm). The null hypothesis was preferred in each case except for Mg and K (analysed by ^{24}Mg and ^{39}K), where the sinusoidal fit with a circadian period was preferred (blue line,

R^2 and fit period \pm s.e.m. are reported). **b**, *Bmal1:luc* bioluminescence data showing no effect of serum concentration on circadian period in U2OS cells in the presence of B-27 supplement, mean \pm s.e.m. ($n = 3$ biological replicates). **c**, Quantification of period and amplitude for data shown in **b** (one-way ANOVA for period, $P = 0.79$; one-way ANOVA for amplitude, $P = 0.01$). **d**, Cellular impedance measurements indicate that U2OS cells do not proliferate upon reaching stationary phase under our assay conditions, reported doubling times (T_d) were calculated from data collected between the dotted lines.



Extended Data Figure 3 | Circadian rhythms of $[\text{Mg}^{2+}]_i$ in *N. crassa* and mouse fibroblasts. **a**, Circadian regulation of $[\text{Mg}^{2+}]_i$ detected by ICP-MS in the fungus *N. crassa* under constant darkness (mean \pm s.e.m., $n = 3$ biological replicates). **b**, Representative (out of three) FRQ immunoblot sampled in parallel. **c**, Quantification of FRQ abundance (mean \pm s.e.m.,

$n = 3$ biological replicates). **d**, Circadian regulation of $[\text{Mg}^{2+}]_i$ measured by luciferase-based assay is dependent upon CRYPTOCHROME in immortalized adult mouse fibroblasts under constant conditions (mean \pm s.e.m., $n = 3$ biological replicates).



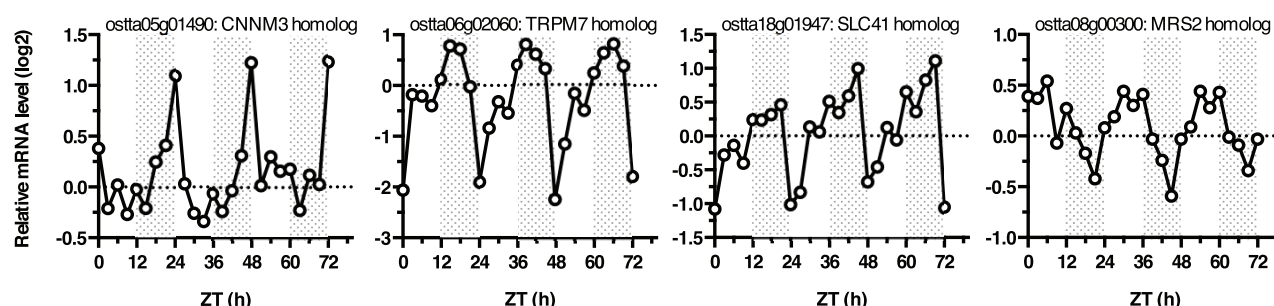
Extended Data Figure 4 | Rhythms of $[Mg^{2+}]_i$ entrain to relevant external cues and are temperature-compensated. **a**, Inversion of 12 h:12 h light/dark entrainment cycles is sufficient to entrain the phase of $[Mg^{2+}]_i$ in *Ostreococcus* cells, measured by luciferase assay under constant light (mean \pm s.e.m., $n = 3$ biological replicates). **b**, From the start of the experiment (S), 3 days of 12 h:12 h temperature cycles between 32 and 37°C, followed by a change to air medium (M), are sufficient to entrain the phase of $[Mg^{2+}]_i$ in U2OS cells measured by ICP-MS over two circadian cycles under constant conditions (mean \pm s.e.m., $n = 3$ biological

replicates). **c**, *Ostreococcus* bioluminescence recordings (CCA1-LUC) at the indicated temperatures ($n = 8$ replicate wells). Vertical dotted lines indicate sampling window for Mg^{2+} assays reported in **d**: assays performed during the second cycle under constant conditions show that circadian $[Mg^{2+}]_i$ rhythms are temperature-compensated ($n = 4$ biological replicates). Each data set was fitted with a Lorentzian curve to estimate peak $[Mg^{2+}]_i$. **e**, No significant difference in temperature compensation (Q_{10}) between CCA1-LUC rhythms and the timing of the second Mg^{2+}_i peak; unpaired t -test P value is reported.

a

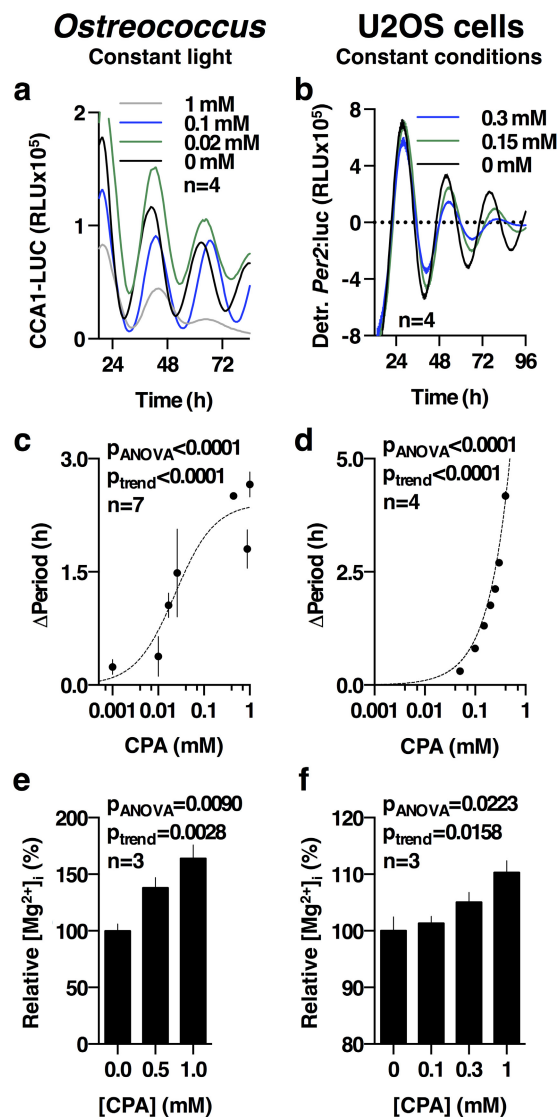
Human gene	Accession (protein)	Protein description	Activity	Membrane	U2OS siRNA knockdown*	Circadian mRNA in mouse tissue†	Closest <i>O. tauri</i> gene‡	<i>O. tauri</i> protein	DELTA-BLAST§	Protein description	Diurnally regulated
TRPM7	NP_001030316	Transient receptor potential cation channel subfamily M member 7, divalent cation channel, alpha kinase family, required for viability	Channel	Plasma membrane	Long period (2/4), Low amplitude (2/4)	Liver, Skeletal Muscle, SCN, Heart	Osta06g02060	XP_003079839	11/25% (7e-34)	Voltage-dependent cation channel	Yes
CNNM3	NP_060093	Cyclin and CBS domain divalent metal cation transport mediator 3, ancient conserved metal transporter domain-containing protein 3	Transporter	Plasma membrane	Long period (3/4)	Kidney, Liver, Colon, White Adipose	Osta05g01490	CEF97914.1	29/47% (5e-59)	RmlC-like jelly roll fold, ion transport domain	Yes
SLC41A1	NP_776253	Solute carrier family 41 member 1, MgtE superfamily	Exchanger	Plasma membrane	Long period (2/4)	Brown Adipose, White Adipose, Lung, Kidney, Heart, SCN, Brain Stem	Osta18g01947	CEG00804	16/33% (3e-9)	SLC41 divalent cation transporters, magnesium transport protein	Yes
MRS2	NP_001273193	Magnesium transporter MRS2 homolog mitochondrial isoform a	Channel	Mitochondrial membrane	Long period (2/4)	Skeletal Muscle	Osta08g00300	CEG01333	20/40% (5e-63)	Mg2+ transporter protein, CorA-like/Zinc transport protein ZntB	Yes
SLC41A2	NP_115524	Solute carrier family 41 member 2, MgtE superfamily	Exchanger	Golgi membrane	Long period (2/4)	Liver, SCN, Macrophages, Cerebellum, Spleen	Osta18g01947	CEG00804	14/30% (5e-13)	SLC41 divalent cation transporters, magnesium transport protein	Yes
MAGT1	NP_115497	Magnesium transporter protein 1	Channel	Plasma membrane	Long period (2/4)	Kidney, Liver, Lung, Heart, Cerebellum, SCN, Adrenal Gland, Hypothalamus, Brain Stem	N/A	N/A	N/A	N/A	N/A

b

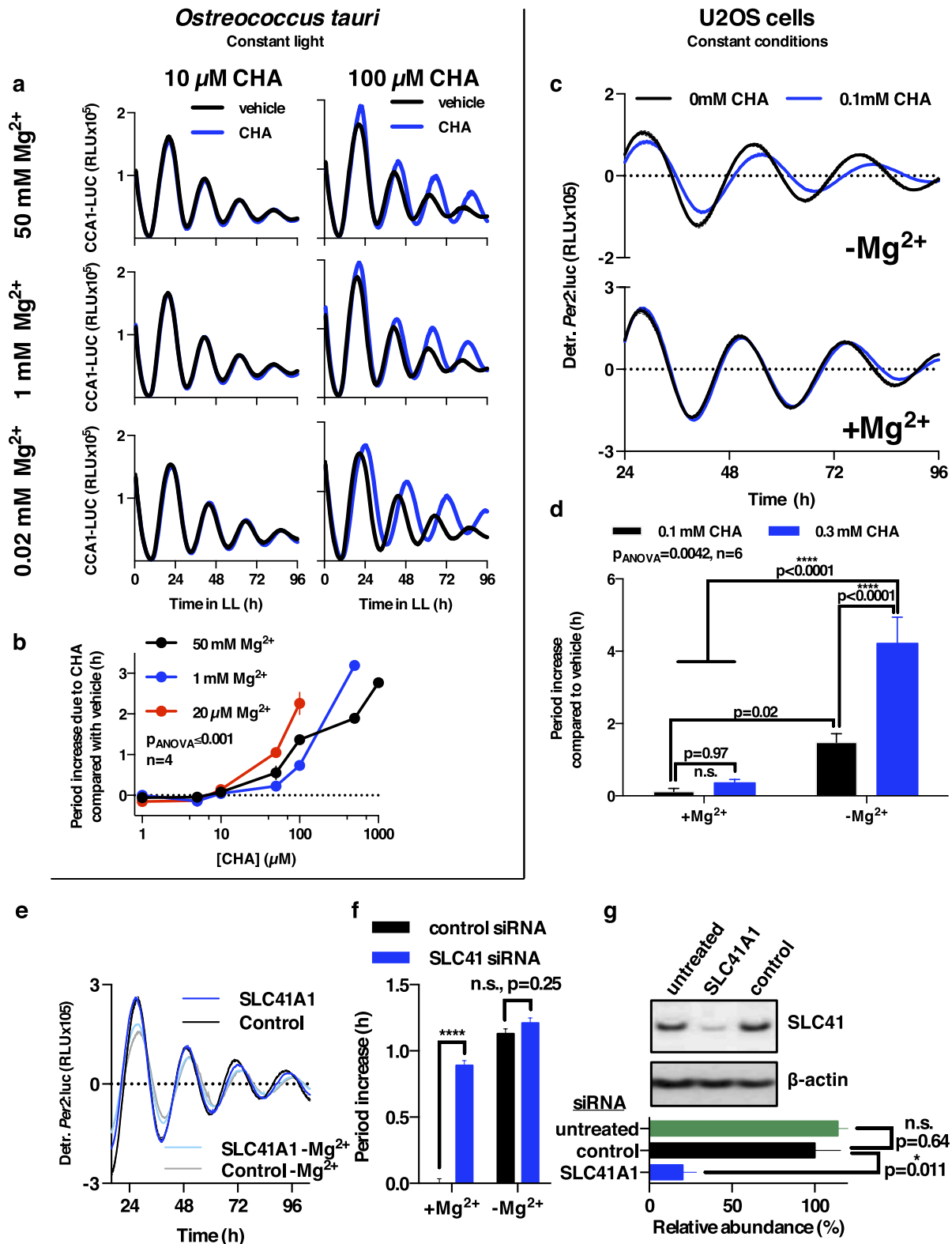


Extended Data Figure 5 | Human magnesium transporters and conservation in *O. tauri*. **a**, Ubiquitously expressed human proteins with a clearly defined Mg^{2+} transport activity⁷ are listed. Note that many additional putative Mg^{2+} -transporters are annotated, with several of these also being circadian-regulated in multiple mouse tissues. **b**, Expression profiles of *Ostreococcus* homologues of mammalian Mg^{2+} channels and

transporters listed in **a**, mined from publicly available microarray data⁴⁵. *From BioGPS^{26,46}. †From CircaDB²⁵ with JTK cycle $P < 0.05$. ‡From the Orcae service^{34,35}. §Percentage sequence identity/similarity with human protein sequence (E value). DELTA-BLAST⁴⁷ performed using default settings. ||From micro-array data⁴⁵ shown in **b**.



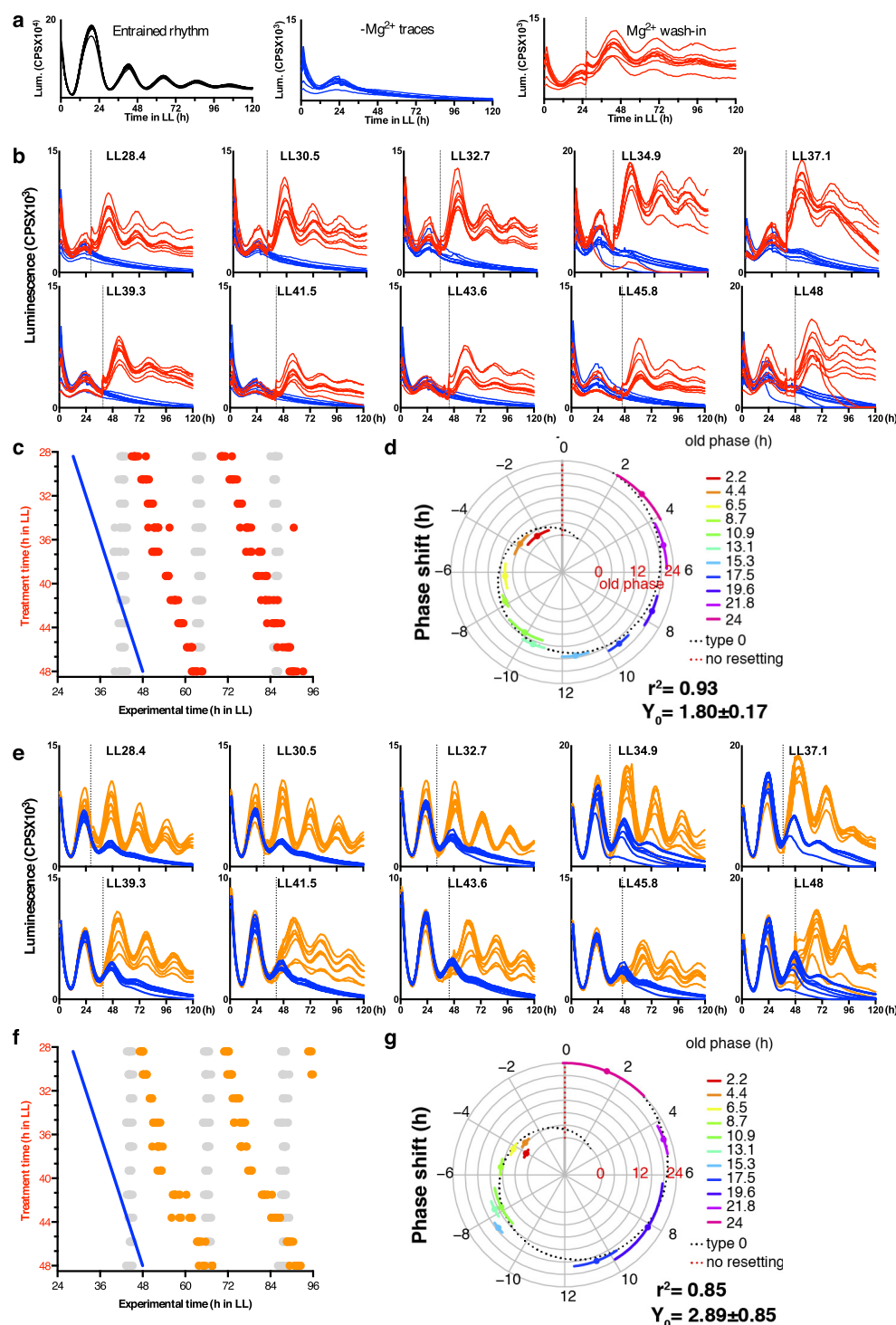
Extended Data Figure 6 | Chronic CPA treatment dose-dependently lengthens period. a–f, Traces (a, b) of the CCA1-LUC (*Ostreococcus*) or *Per2:luc* (U2OS cells) reporters, showing the effect of inhibition of magnesium transport by $Co(NH_3)_5Cl^{2+}$ (CPA) upon period dose–response (c, d) and upon $[Mg^{2+}]_i$ (e, f). All plots show mean \pm s.e.m., with biological replicate numbers (*n*) indicated. *P* values report significance by one-way ANOVA and post-test for linear trend.



Extended Data Figure 7 | Period lengthening by CHA and SLC41 knockdown is dependent upon extracellular magnesium.

a, Extracellular magnesium-depletion and CHA act synergistically to lengthen circadian period in *Ostreococcus* cells (mean \pm s.e.m., $n=4$ biological replicates). **b**, Quantification of period lengthening by CHA at different concentrations of extracellular magnesium (mean \pm s.e.m., $n=4$ biological replicates). P value for two-way ANOVA (interaction effect) is reported. **c**, Extracellular magnesium-depletion and CHA act synergistically to lengthen circadian period in human U2OS cells (mean \pm s.e.m., $n=6$ biological replicates). **d**, Quantification of period lengthening by CHA in Mg^{2+} -depleted versus normal media

(mean \pm s.e.m., $n=4$ biological replicates). P values for two-way ANOVA (interaction effect) and Fisher's exact test are reported. **e**, Period lengthening due to knockdown of plasma membrane $\text{Mg}^{2+}/\text{Na}^{+}$ antiporter SLC41A1 is attenuated by depletion of extracellular magnesium (mean \pm s.e.m., $n=8$ biological replicates). **f**, Quantification of period lengthening due to knockdown of SLC41A1 in normal versus Mg^{2+} -depleted media (mean \pm s.e.m., $n=8$ biological replicates); two-way ANOVA interaction effect, $P < 0.0001$. P values for Sidak's multiple comparisons test are also reported. **g**, Quantification of SLC41A1 knockdown efficacy. Unpaired t -test P values are reported, a representative immunoblot (of three) is shown (mean \pm s.e.m., $n=3$ biological replicates).

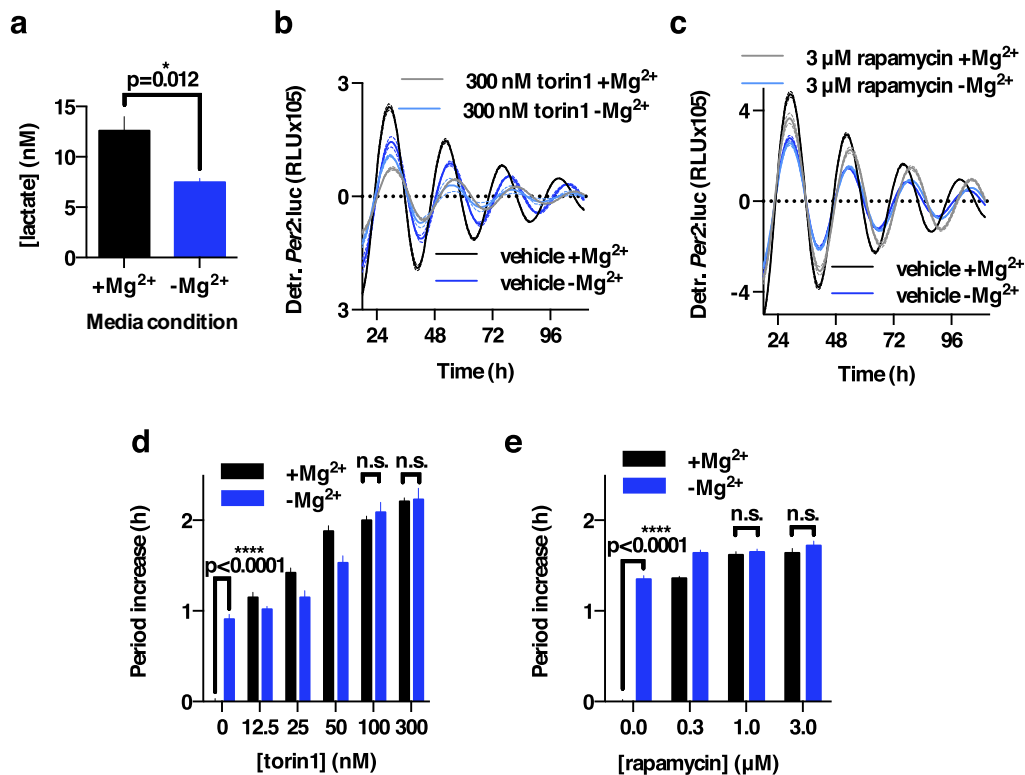


Extended Data Figure 8 | Bioluminescence data of wedge experiment. Peak expression phase of the clock protein CCA1 was analysed upon re-introduction of magnesium to cultures in low extracellular magnesium, to test whether the phase of cellular rhythms is dictated by the prior phase of entrainment or by this enforced transition from low to high $[Mg^{2+}]_i$. **a**, Bioluminescence traces showing that circadian rhythms in *Ostreococcus* are reversibly attenuated by depletion of extracellular Mg^{2+} , and restored by Mg^{2+} wash-in. **b**, **e**, Bioluminescence traces from cells in low extracellular magnesium (**b**; $5 \mu M$, **e**; $20 \mu M$) with rhythms rescued by release into media containing normal physiological concentrations of magnesium at the indicated times (vertical dotted lines) in constant light (LL), compared with their respective controls where no magnesium was added in (blue traces). Data from seven or eight replicate wells are shown in each panel. **c**, **f**, Summary graphs where results from **b**, **e** are plotted in circadian wedge graphs: peak phases of CCA1-LUC rhythms

in untreated control cells (grey dots) are compared with peak phase of rhythms reinstated by introduction of physiological magnesium following depletion to $5 \mu M$ (**c**, red dots) or $20 \mu M$ (**f**, orange dots), revealing that the phase of resulting rhythms is dictated solely by the phase of magnesium reintroduction (blue line). **d**, **g**, Radial plots of phase shift (mean \pm s.d., circumferential axis) depicted in **c** and **f**, versus phase before addition of Mg^{2+} to normal levels (old phase, radial axis and colour). The expected phase responses for type 0 resetting (black dotted line) and no resetting (red dotted line) are indicated. The goodness of fit (R^2) and y intercept (Y_0) to the type 0 model are shown. Dose-dependent effects of intracellular magnesium on a critical clock parameter are confirmed by the observation that resetting was less strong when magnesium was reintroduced to cells adapted to intermediate levels of extracellular magnesium (**e–g**) compared with lowest extracellular magnesium (**b–d**).

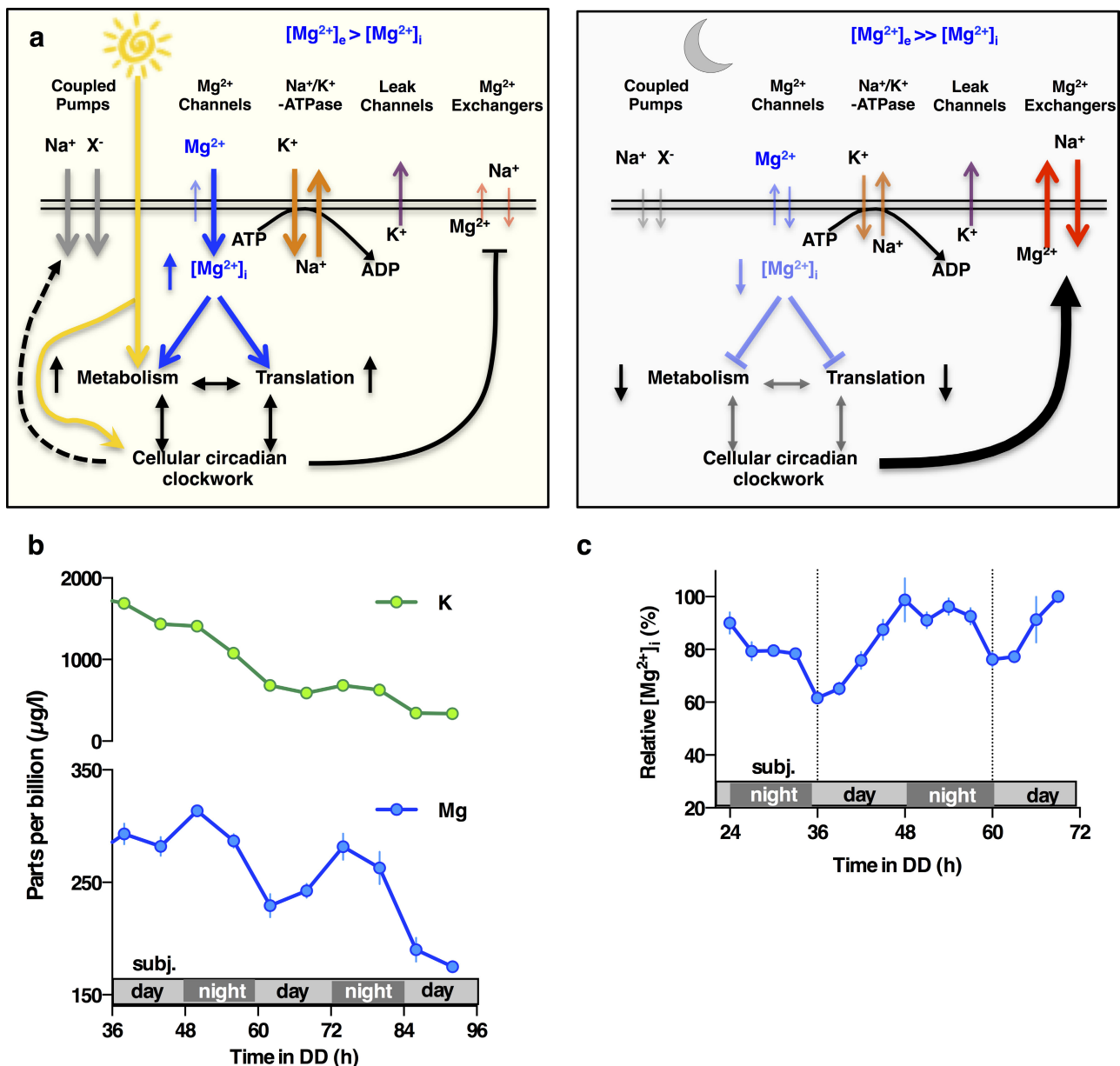
U2OS cells

Constant conditions



Extended Data Figure 9 | The effects of magnesium depletion and role of mTOR. **a**, Extracellular lactate was measured in U2OS cells after 24 h in Mg²⁺-depleted compared with normal media. **b**, **c**, Combined action of extracellular magnesium depletion and mTOR inhibition using torin1 (**b**, $n = 3$ biological replicates) or rapamycin (**c**, $n = 6$ biological replicates) to lengthen circadian period in U2OS cells is less than additive (mean \pm s.e.m.). **d**, **e**, Quantification of period lengthening due to

torin1 (**d**, $n = 3$) and rapamycin (**e**, $n = 6$) in Mg²⁺-depleted compared with normal media (mean \pm s.e.m.). Note the apparent 'ceiling effect' at high concentrations of both drugs, such that Mg²⁺ depletion elicits no additional lengthening of cellular circadian period. Two-way ANOVA interaction effect: $P < 0.0001$ for both drugs versus Mg²⁺. Selected P values for Sidak's multiple comparisons test are also reported (NS, $P > 0.33$).



Extended Data Figure 10 | Factors potentially contributing to maintenance of membrane electroneutrality in light of $[Mg^{2+}]_i$ oscillations. **a**, Model indicating potential ion fluxes that might explain how clock-regulated $[Mg^{2+}]_i$ oscillations impact on global cellular metabolism while membrane electroneutrality is maintained, during the day versus the night. The observed phase dependency of acute CHA was different between *Ostreococcus* and U2OS cells (Fig. 4a–c), and is consistent with the very different environmental niches inhabited by a marine alga compared with a peripheral human tissue. In *Ostreococcus*, CHA maintained $[Mg^{2+}]_i$ at daytime levels when added before the normal trough, resulting in increased night-time translation and a concomitant reduction in relative ATP levels. This result suggests that *Ostreococcus* pumps magnesium out of the cell during the dark period, against a large electrochemical potential gradient (magnesium is the second most abundant cation in seawater, at 50 mM in this study) to globally

down-tune ATP turnover. In U2OS cells, CHA treatment significantly reduced $[Mg^{2+}]_i$ accumulation and translation rates as well as significantly increasing ATP levels when added before the $[Mg^{2+}]_i$ peak. Human cells inhabit an environment where nutrient availability is homeostatically regulated (0.8 mM magnesium in cell culture medium). As such, circadian regulation of increased magnesium transport into the cell during the feeding, active phase of day serves to facilitate higher metabolic rate constants. Note that light has no direct effect on the clock in human peripheral cells, instead being mediated by systemic cues. **b**, **c**, $[Mg^{2+}]_i$ oscillations persist in transcriptionally inactive *Ostreococcus* cells kept in constant darkness, as analysed both by ICP-MS (**b**) and by luciferase assay (**c**), indicating that circadian regulation of ion transport can occur post-translationally in addition to its transcriptional regulation (mean \pm s.e.m., $n = 3$ for ICP-MS data and $n = 4$ for luciferase assays (biological replicates)).

Age-dependent modulation of vascular niches for haematopoietic stem cells

Anjali P. Kusumbe¹, Saravana K. Ramasamy¹, Tomer Itkin², Maarja Andaloussi Mäe³, Urs H. Langen¹, Christer Betsholtz^{3,4}, Tsvee Lapidot² & Ralf H. Adams¹

Blood vessels define local microenvironments in the skeletal system, play crucial roles in osteogenesis and provide niches for haematopoietic stem cells^{1–6}. The properties of niche-forming vessels and their changes in the ageing organism remain incompletely understood. Here we show that Notch signalling in endothelial cells leads to the expansion of haematopoietic stem cell niches in bone, which involves increases in CD31-positive capillaries and platelet-derived growth factor receptor- β (PDGFR β)-positive perivascular cells, arteriole formation and elevated levels of cellular stem cell factor. Although endothelial hypoxia-inducible factor signalling promotes some of these changes, it fails to enhance vascular niche function because of a lack of arterIALIZATION and expansion of PDGFR β -positive cells. In ageing mice, niche-forming vessels in the skeletal system are strongly reduced but can be restored by activation of endothelial Notch signalling. These findings indicate that vascular niches for haematopoietic stem cells are part of complex, age-dependent microenvironments involving multiple cell populations and vessel subtypes.

Different vessel subtypes have distinct roles in the skeletal system. Veins drain the bone marrow (BM) cavity, while arteries deliver oxygen-rich blood and are thought to provide niches for quiescent haematopoietic stem cells (HSCs)⁷. Type H capillaries, characterized by high expression of CD31 (CD31^{hi}) and endomucin (Emcn^{hi}), connect to arterioles (Extended Data Fig. 1a, b), are surrounded by osteoprogenitors and release factors promoting osteogenesis^{4,6}. By contrast, type L (CD31^{lo} Emcn^{lo}) vessels, which correspond to BM sinusoids, lack arteriolar connections and osteoprogenitor association⁴. For deeper characterization of vessel subpopulations, we analysed bones from *Efnb2*^{GFP/+} knock-in⁸ mice (Extended Data Fig. 1c), which express green fluorescent protein (GFP) under control of the gene encoding ephrin-B2, an arterial marker⁹. *Efnb2*^{GFP/+} signals labelled arteries and distal arterioles, both of which lack Emcn expression, and the adjacent Emcn⁺ type H endothelial cells (ECs) in metaphysis and endosteum (Extended Data Fig. 2a–c). Quantitative PCR (qPCR) showed significantly higher expression of *Efnb2* and *Sox17*, a transcription factor and regulator of arterial differentiation¹⁰, in sorted type H ECs relative to type L ECs (Extended Data Fig. 2d). Immunostaining confirmed expression of Sox17, ephrin-B2 and Neuropilin-1 (Nrp1), another marker of arterial ECs, in arteries and type H capillaries (Fig. 1a and Extended Data Fig. 3a–d). Consistent with previous reports², bone arteries also showed Sca-1 (Ly-6A/E) immunostaining, which extended to type H endothelium (Extended Data Fig. 3e, f). Proliferation, a key feature of angiogenic vessel growth^{11,12}, was prominent in type H ECs and distal arterioles, whereas EdU labelling was negligible in α -smooth muscle actin (α -SMA)-covered arteries (Extended Data Fig. 3g–i). Thus, arterioles and type H vessels are found in direct proximity and share many features distinguishing them from BM sinusoidal (type L) capillaries.

In addition to BM ECs, important roles in the regulation of HSCs have been attributed to the mesenchymal lineage and, in particular, Nestin⁺ mesenchymal stem cells (MSCs)¹³. Cells expressing PDGFR β , the receptor for platelet-derived growth factor B (PDGF-B) and a marker of pericytes and other mesenchymal cell types¹⁴, were primarily located around type H capillaries and arteries but not type L vessels (Fig. 1b and Extended Data Fig. 4a). Perivascular mesenchymal cells

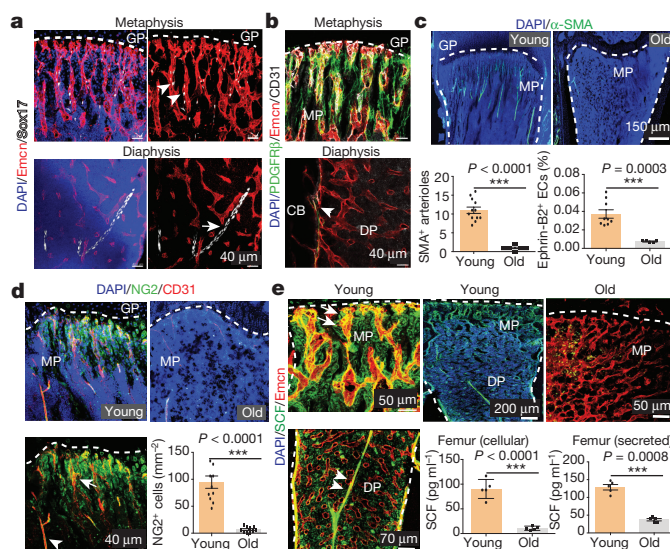


Figure 1 | Bone vessel subtypes and properties. **a**, Confocal images showing Emcn (red) and Sox17 (white) in tibial metaphysis or diaphysis. Nuclei, 4',6-diamidino-2-phenylindole (DAPI; blue). Arrow, artery; arrowhead, type H capillaries. GP, growth plate. **b**, PDGFR β ⁺ cells (green) surround arteries and Emcn⁺ (red) CD31⁺ (white) vessels in metaphysis (MP) and endosteum (arrowhead) but not diaphyseal (DP) sinusoidal vessels in 4-week-old tibia. CB, compact bone. **c**, α -SMA⁺ arterioles in 5- (young) and 65-week-old (old) tibial metaphysis. Nuclei, DAPI (blue). Dashed lines mark growth plate and compact bone. Graphs show quantitative analysis of α -SMA⁺ metaphyseal arteries ($n = 12$, two independent experiments) and flow cytometric analysis of ephrin-B2⁺ ECs ($n = 9$ young and 5 old mice). Data represent mean \pm s.e.m. P values, two-tailed unpaired t -test. **d**, NG2⁺ (green) cells and CD31⁺ (red) vessels in sections from 5- (young, left) and 65-week-old (old) tibial metaphysis. Graph shows quantification of NG2⁺ mesenchymal cells excluding NG2⁺ bone cells. Data represent mean \pm s.e.m. (two tibiae from five mice). P values, two-tailed unpaired t -test. **e**, Confocal images showing SCF (green) staining in arteries (arrowheads) and Emcn⁺ (red) metaphyseal type H vessels (arrow) of young (3-week-old) but not old (65-week-old) tibiae. Quantitative analysis of cellular (lysate of washed total BM cells) and extracellular SCF (bone supernatant) in young/old long bone by ELISA. Data represent mean \pm s.d. ($n = 5$). P values, two-tailed unpaired t -test.

¹Max-Planck-Institute for Molecular Biomedicine, Department of Tissue Morphogenesis, and University of Münster, Faculty of Medicine, D-48149 Münster, Germany. ²Department of Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel. ³Vascular Biology Program, Department of Immunology, Genetics and Pathology, Uppsala University, SE-751 85 Uppsala, Sweden. ⁴Department of Medical Biochemistry and Biophysics, Division of Vascular Biology, Karolinska Institute, Scheeles väg 2, SE-171 77 Stockholm, Sweden.

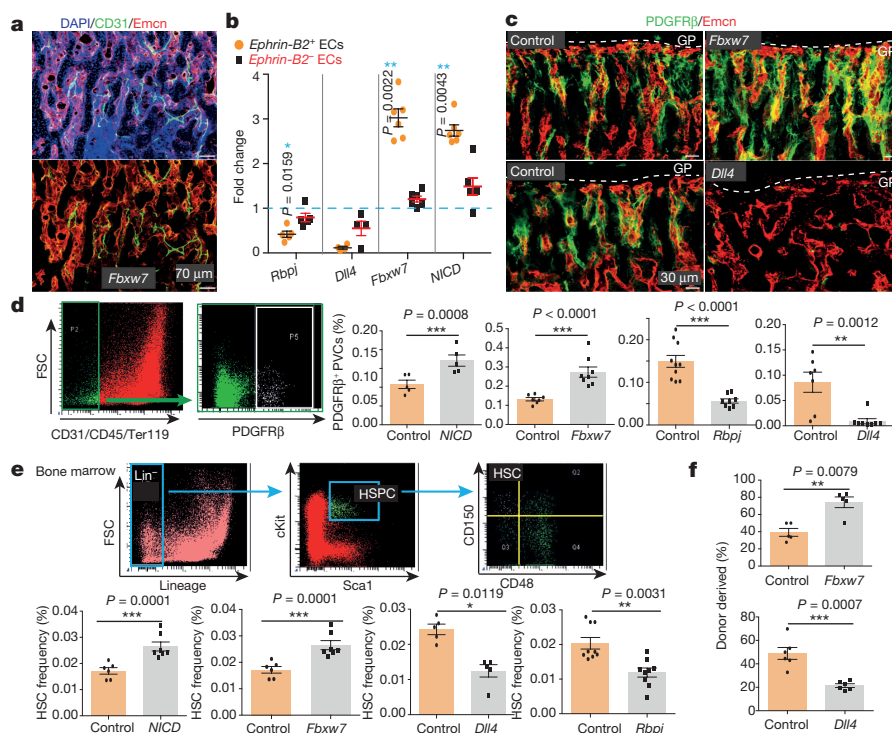


Figure 2 | Endothelial Notch regulates BM HSC numbers. **a**, Formation of numerous CD31⁺ (green) and Emcn⁺ (red) arterioles in 4-week-old *Fbxw7*^{ΔEC} tibial diaphysis. Nuclei, DAPI (blue). **b**, Quantitative analysis of ephrin-B2⁺ and ephrin-B2[−] ECs in *NICD*^{ΔEC} (*n* = 6 mutants and 5 controls), *Fbxw7*^{ΔEC} (*n* = 6 and 5), *Rbpj*^{ΔEC} (*n* = 5) and *Dll4*^{ΔEC} (*n* = 4 and 5) long bone relative to Cre-negative controls (dotted line). Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. Total ECs were identified as CD45[−] Ter119[−] CD31⁺ cells. **c**, Representative images of *Fbxw7*^{ΔEC}, *Dll4*^{ΔEC} or control tibial metaphysis showing PDGFRβ⁺ (green) cells around Emcn⁺ (red) capillaries. **d**, Flow cytometric quantification of CD31[−] CD45[−] Ter119[−] PDGFRβ⁺ cells, as shown in representative dot plots. Note increase in PDGFRβ⁺ cells in Notch

gain-of-function bone (*NICD* *n* = 5; *Fbxw7* *n* = 8 and 7) but decrease in *Rbpj*^{ΔEC} (*n* = 8 and 10) and *Dll4*^{ΔEC} (*n* = 7) mutants. Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. **e**, Flow cytometric quantification (as shown in dot plots) of HSCs in *NICD*^{ΔEC} (*n* = 7 and 6), *Fbxw7*^{ΔEC} (*n* = 8 and 10), *Rbpj*^{ΔEC} (*n* = 8 and 10) and *Dll4*^{ΔEC} (*n* = 7) long bones relative to Cre-negative littermates. Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. **f**, Analysis of long-term repopulating (LTR)-HSC contribution of *Fbxw7*^{ΔEC} and *Dll4*^{ΔEC} donor-derived BM cells, as determined by flow cytometry 7 months after competitive transplantation together with recipient-derived CD45.1 BM cells into lethally irradiated mice. Data represent mean ± s.e.m. (*n* = 5 donors for *Fbxw7*^{ΔEC}, *n* = 6 donors for *Dll4*^{ΔEC}).

also expressed NG2 (Extended Data Fig. 4b–d). Similar to the localization of type H capillaries, distal arterioles and PDGFRβ⁺ NG2⁺ mesenchymal cells, CD150⁺ CD48[−] Lin[−] Sca1⁺ c-Kit⁺ HSCs were more abundant in the dissected metaphysis than in BM flushed out from the diaphysis (Extended Data Fig. 4e). In addition, CD150⁺ CD48[−] cells were found in proximity to type H vasculature in thick cryosections (100 μm) of the metaphysis and endosteum (Extended Data Fig. 4f–i).

Ageing is associated with changes in HSC number and quality^{15,16}, reduced skeletal blood flow¹⁷ and alterations in BM stroma¹⁸. We have previously shown that type H ECs and associated osteoprogenitors decline significantly in adult and aged mice⁴. As type H capillaries connect to arterioles⁶, we compared the abundance of arteries in young (4-week-old) and aged (65- to 70-week-old) bone. Strikingly, aged tibiae contained fewer α-SMA-covered arteries and *Efnb2*⁺ ECs (Fig. 1c and Extended Data Fig. 5a, b). Likewise, the abundance of PDGFRβ⁺ or NG2⁺ perivascular cells was reduced in aged long bone (Fig. 1d and Extended Data Fig. 5d, e). Stem cell factor (SCF) is a cytokine with crucial roles in HSC homing and maintenance^{1,19}. CD31^{hi} Emcn^{hi} ECs in the metaphysis and endosteum as well as vessel-associated cells covering type H capillaries and arteries were strongly positive for SCF, whereas expression in diaphyseal type L vasculature was weaker (Fig. 1e). Accordingly, expression of the *Kitl* gene encoding murine SCF was significantly higher in freshly isolated type H relative to type L ECs (Extended Data Fig. 5c). Consistent with the reduction of arterioles, type H capillaries and PDGFRβ⁺/NG2⁺ perivascular cells, SCF levels also declined in aged mice (Fig. 1e).

The Notch pathway promotes artery formation, and the abundance of small calibre arterioles and type H ECs was strongly increased in EC-specific Notch gain-of-function mutant bone⁶. Analysis of Notch1 activity in bone from *NICD*-Cre knock-in mice²⁰ in the *Rosa26-mT/mG*²¹ Cre reporter background showed recombination (GFP signal) in type H ECs, perivascular cells in metaphysis and endosteum, and in arteries (Extended Data Fig. 5f–h). EC-specific overexpression of the Notch1 intracellular domain (*NICD*^{ΔEC}) or inactivation of the gene encoding Fbxw7 (*Fbxw7*^{ΔEC}), which mediates polyubiquitination and thereby proteasomal degradation of active Notch²², induced arteriole formation and expansion of Sca1⁺ ECs and ephrin-B2⁺ ECs (Fig. 2a, b and Extended Data Fig. 6a, b, d). This was accompanied by increases in α-SMA coverage, PDGFRβ⁺ perivascular cells, MSC frequency, SCF levels and HSC frequency in *Fbxw7*^{ΔEC} BM (Fig. 2c–e and Extended Data Figs 6c–e and 7a–d). The increase in HSC frequency was confirmed by primary and secondary competitive reconstitution assays of irradiated mice transplanted with *Fbxw7*^{ΔEC} or control BM cells (Fig. 2f and Extended Data Fig. 7e, f). The frequencies of different haematopoietic lineages remained unaltered in *Fbxw7*^{ΔEC} BM (Extended Data Fig. 7g). These data establish that manipulation of the endothelium can enhance vascular niche function leading to increased HSC frequency.

Notch signalling in the endothelium requires the Notch ligand Dll4 and the DNA-binding protein RBPJ, which controls gene expression downstream of activated Notch⁶. The abundance of arteries, the number of ephrin-B2⁺ ECs and PDGFRβ⁺ cells, SCF levels and HSC frequency were all reduced in EC-specific *Rbpj*^{ΔEC} or *Dll4*^{ΔEC}

Notch loss-of-function long bone (Fig. 2b–e and Extended Data Fig. 7c, d). *Dll4*^{ΔEC} BM cells also showed reduced long-term repopulation activity in competitive reconstitution assays (Fig. 2f). Notch1 is the main Notch receptor controlling EC behaviour²³. Consistently, vascular organization, abundance of type H ECs and frequency of CD150⁺ CD48⁺ Lin⁺ Sca1⁺ c-Kit⁺ HSCs were not altered in EC-specific *Notch2*^{ΔEC} or global *Notch4* mutant bone (Extended Data Fig. 7h–j). Notch3 expression was not detectable in bone ECs (Extended Data Fig. 7k).

Similar to Notch, the hypoxia-inducible factor (HIF) pathway positively controls type H EC and osteoprogenitor abundance⁴. The metabolic environment, oxygenation (pimonidazole staining) and expression of *Hif1a* and *Epas1* (*Hif2a*) in bone change during ageing (Extended Data Fig. 8a, b). In addition to its oxygen-dependent degradation, the activity and expression of HIF-1α are promoted by growth factors and cytokines via activation of the phosphatidylinositol 3-kinase (PI3K) or mitogen-activated protein kinase (MAPK) pathways²⁴. Expression of several growth factors and cytokines as well as levels of the phosphorylated MAPK extracellular signal-regulated kinase (ERK) were reduced in aged metaphysis (Extended Data Fig. 8c, d), which might explain the age-dependent reduction in endothelial *Hif1a*. The von Hippel–Lindau (VHL) protein is involved in the degradation of HIF, and EC-specific inactivation of the *Vhl* gene led to a striking expansion of type H vasculature and perivascular osteoprogenitors⁴. Highlighting the essential role of HIF-1α in these VHL-mediated effects, loss of type H ECs and reduced expression of EC-derived growth factors were not reversed in *Hif1a*^{ΔEC} *Vhl*^{ΔEC} double mutants (Extended Data Fig. 8e–g). In addition to decreased type H ECs, *Hif1a*^{ΔEC} mutants exhibited significant reductions in Sca1⁺ ECs, ephrin-B2⁺ ECs, PDGFRβ⁺ perivascular cells, SCF levels and HSC frequency (Fig. 3a–f and Extended Data Fig. 8h–j).

Arterial development frequently involves the incorporation of ECs from surrounding capillaries^{25,26}. As larger arteries were devoid of HIF-1α signal (Extended Data Fig. 9a), *Hif1a*^{ΔEC} arterial defects might be caused by alterations in adjacent type H capillaries. Indeed, despite increased type H ECs in *Vhl*^{ΔEC} mice, these mutants did not show significant increases in morphologically identifiable arterioles, Sca1⁺ ephrin-B2⁺ ECs and PDGFRβ⁺ perivascular cells (Fig. 3c, e and Extended Data Figs 8k, l and 9b, c). Although secreted SCF levels were elevated, cellular SCF levels were not significantly altered and HSC frequency was not increased in *Vhl*^{ΔEC} mutants (Fig. 3d, f). These results are consistent with previous studies demonstrating membrane-bound SCF as a more potent stimulator of c-Kit and effective driver of HSC adhesion in BM relative to secreted SCF^{27,28}. In cultured primary bone ECs and PDGFRβ⁺ cells, HIF stabilization induced by deferoxamine mesylate (DFM) significantly increased secreted but not cellular/membrane-bound SCF (Extended Data Fig. 9d, e).

To further dissect the differences and potential interplay between endothelial Notch and HIF signalling, we generated *Hif1a*^{ΔEC} *NICD*^{ΔEC}, *Rbpj*^{ΔEC} *Vhl*^{ΔEC} and *NICD*^{ΔEC} *Vhl*^{ΔEC} double-mutant mice. Notch activation in the absence of HIF-1α (*Hif1a*^{ΔEC} *NICD*^{ΔEC}) and stabilization of HIF-1 in the absence of Notch signalling (*Rbpj*^{ΔEC} *Vhl*^{ΔEC}) independently mediated the expansion of type H ECs (Extended Data Fig. 9f). Analysis of double mutants also corroborated Notch signalling as being essential for expansion of PDGFRβ⁺ perivascular cells, elevation of cellular SCF levels and enhancement of HSC numbers (Extended Data Fig. 9g–i). An important role of perivascular cells was further supported by findings in *Pdgfrb*^{ΔEC} mice overexpressing PDGF-B in ECs. *Pdgfrb*^{ΔEC} tibiae contained more PDGFRβ⁺ and α-SMA⁺ cells and had higher levels of cellular SCF and a higher frequency of HSCs (Fig. 3g–j). Taken together, endothelial Notch but not HIF signalling can enhance the frequency of HSCs by improving key aspects of vascular niche function—namely arteries, type H ECs, perivascular cells and cellular SCF production. These effects are distinct from the previously reported role of the Notch ligand Jagged1 in ECs, which promotes haematopoiesis through Notch activation in

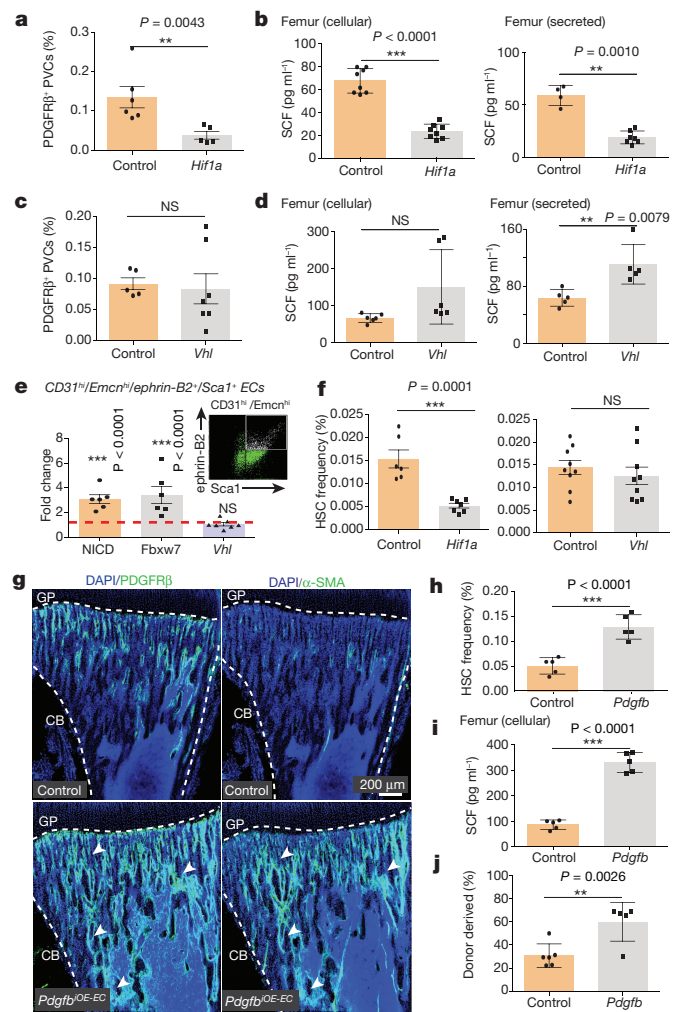


Figure 3 | Endothelial HIF signalling and vascular niches.

a, Flow cytometric analysis of CD31⁺ CD45⁺ Ter119⁺ PDGFRβ⁺ cells in *Hif1a*^{ΔEC} and control long bone. Data represent mean ± s.e.m. (*n* = 5 mutants and 6 controls). *P* values, two-tailed unpaired *t*-test. **b**, ELISA analysis of cellular and secreted SCF levels in *Hif1a*^{ΔEC} and control long bone. Data represent mean ± s.d. *P* values, two-tailed unpaired *t*-test. **c**, Flow cytometric quantification of CD31⁺ CD45⁺ Ter119⁺ PDGFRβ⁺ cells from *Vhl*^{ΔEC} and control long bone. Data represent mean ± s.e.m. (*n* = 7 and 5). **d**, Graphs showing ELISA analysis of cellular and secreted SCF levels in *Vhl*^{ΔEC} long bone relative to controls. Data represent mean ± s.d. *P* values, two-tailed unpaired *t*-test. **e**, Fold change in frequency of ephrin-B2⁺ Sca1⁺ ECs in *NICD*^{ΔEC} (*n* = 6) and *Fbxw7*^{ΔEC} (*n* = 6) but not in *Vhl*^{ΔEC} (*n* = 7) long bone relative to littermate controls, as determined by flow cytometric analysis (see representative dot plot). Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. **f**, Flow cytometric analysis of HSCs in *Hif1a*^{ΔEC} (*n* = 7 and 6) and *Vhl*^{ΔEC} (*n* = 9) long bone relative to Cre-negative littermates. Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. **g**, Representative tile scan confocal images showing PDGFRβ (green, left) or α-SMA (green, right) in *Pdgfrb*^{ΔEC} and control tibia sections. Nuclei, DAPI (blue). Dashed lines mark growth plate or cortical bone. **h**, Flow cytometric quantitation of HSCs in *Pdgfrb*^{ΔEC} long bone relative to Cre-negative littermates (*n* = 5). Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. **i**, ELISA analysis of cellular SCF levels in *Pdgfrb*^{ΔEC} long bone. Data represent mean ± s.d. *P* values, two-tailed unpaired *t*-test. **j**, Analysis of HSC contribution of *Pdgfrb*^{ΔEC} donor-derived BM cells, as determined by flow cytometry at 14 weeks after competitive transplantation together with recipient-derived CD45.1 BM cells into lethally irradiated mice. Data represent mean ± s.e.m. (*n* = 5 and 6 donors).

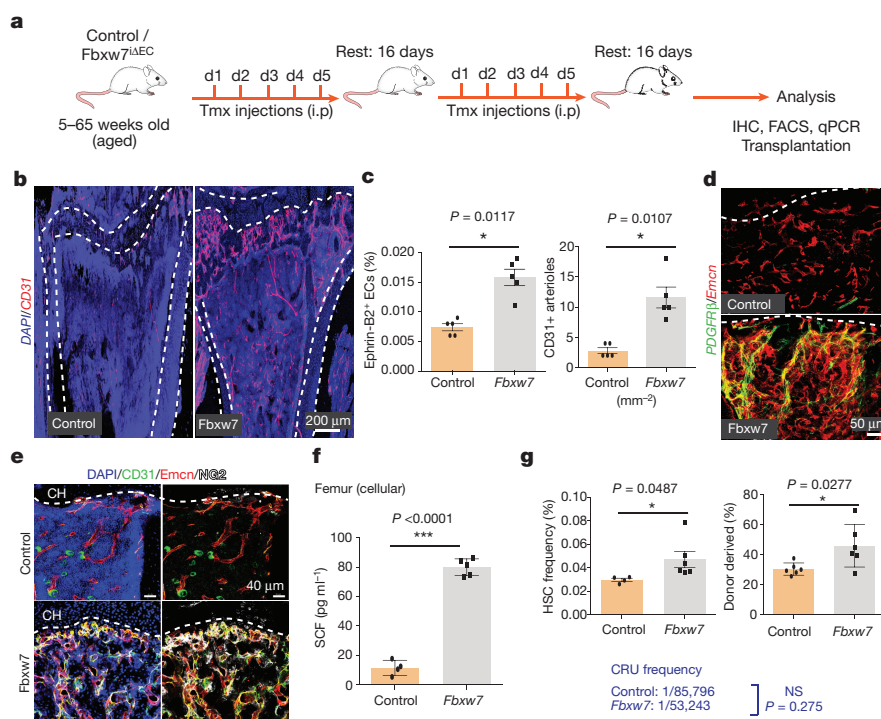


Figure 4 | Endothelial Notch reactivates HSC niches in aged mice.

a, Tamoxifen (Tmx) intraperitoneal (i.p.) administration strategy for aged *Fbxw7*^{ΔEC} mutants. Injection days (d1–d5) and rest periods are indicated. **b**, Increase in CD31 (red) immunostained vessels in aged *Fbxw7*^{ΔEC} tibia. Nuclei, DAPI (blue). **c**, Quantification of CD45⁺ Ter119⁺ CD31⁺ ephrin-B2⁺ ECs (by flow cytometry) and of CD31⁺ arterioles (by morphology) in *Fbxw7*^{ΔEC} and control long bone. Data represent mean ± s.e.m. (*n* = 5). *P* values, two-tailed unpaired *t*-test. **d**, Increase of PDGFRβ⁺ (green) cells associated with Emcn⁺ (red) vessels in *Fbxw7*^{ΔEC} tibia relative to control. Nuclei, DAPI (blue). **e**, Increase in

NG2⁺ (white) cells and abundance of CD31⁺ (green) and Emcn⁺ (red) vessels in *Fbxw7*^{ΔEC} tibia. Dashed lines mark the chondrocyte zone (CH). **f**, ELISA analysis of cellular SCF in *Fbxw7*^{ΔEC} and control long bone. Data represent mean ± s.d. (*n* = 5 and 4), *P* values, two-tailed unpaired *t*-test. **g**, Flow cytometric quantification (left) of HSCs in aged *Fbxw7*^{ΔEC} (*n* = 6 and 4). Data represent mean ± s.e.m. *P* values, two-tailed unpaired *t*-test. Analysis of HSC contribution of BM cells from aged *Fbxw7*^{ΔEC} and control donors, as determined by flow cytometry at 16 weeks after competitive transplantation with age-matched CD45.1 BM cells into lethally irradiated recipients. Data represent mean ± s.e.m. (*n* = 6).

haematopoietic stem/progenitor cells without detectable alterations in the BM vasculature²⁹.

Given the decline of arteries, type H capillaries and SCF levels during ageing, we next addressed whether Notch activation in aged mice would improve the niche properties of bone endothelium (Fig. 4a). Remarkably, aged *Fbxw7*^{ΔEC} bones showed profound expansion of CD31^{hi} capillaries and CD31⁺ arterioles (Fig. 4b, c). The numbers of ephrin-B2⁺ ECs and of PDGFRβ⁺ or NG2⁺ perivascular cells, SCF levels and HSC frequency were significantly increased in mutants (Fig. 4c–g and Extended Data Fig. 10a), establishing that vascular niche function can be enhanced in aged mice. Ageing is associated with an accumulation of HSCs, which display cell-intrinsic alterations such as DNA damage and reduced functionality^{15,30}. Consistent with these studies, increase in the donor-derived chimaerism could not be achieved upon transplantation of aged *Fbxw7*^{ΔEC} BM cells along with young competitor BM (Extended Data Fig. 10b). Limiting dilution analysis, which allows the quantification of functional HSCs, indicated no statistically significant increase upon transplantation of aged *Fbxw7*^{ΔEC} BM cells relative to control (Extended Data Fig. 10c, d). Analysis of DNA damage by γ-H2AX immunostaining also confirmed the persistence of age-related cell intrinsic impairments in *Fbxw7*^{ΔEC} HSCs (Extended Data Fig. 10e). Thus, changes in the bone vasculature of old mice can boost the number of HSCs and progenitor cells but cannot restore the functionality of aged HSCs because of the persistence of cell-autonomous alterations.

Our results provide novel insight into the age-dependent changes occurring in the HSC niche. We propose that the enhancement of functional vascular niches in bone requires type H ECs, the formation of arterioles/arteries and the expansion of perivascular cells.

Thus, vascular HSC niches are complex microenvironments involving multiple different cell populations and vessel subtypes. Our findings also establish that manipulation of the endothelium is sufficient for the improvement of vascular HSC niche function, which suggests the existence of molecular pathways coupling the behaviour of ECs and perivascular mesenchymal cells in bone.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 January 2015; accepted 3 March 2016.

Published online 13 April 2016.

- Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
- Hooper, A. T. et al. Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4**, 263–274 (2009).
- Kiel, M. J. et al. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Kusumbe, A. P., Ramasamy, S. K. & Adams, R. H. Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone. *Nature* **507**, 323–328 (2014).
- Acar, M. et al. Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature* **526**, 126–130 (2015).
- Ramasamy, S. K., Kusumbe, A. P., Wang, L. & Adams, R. H. Endothelial Notch activity promotes angiogenesis and osteogenesis in bone. *Nature* **507**, 376–380 (2014).
- Kunisaki, Y. et al. Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* **502**, 637–643 (2013).
- Davy, A., Bush, J. O. & Soriano, P. Inhibition of gap junction communication at ectopic Eph/ephrin boundaries underlies craniofrontonasal syndrome. *PLoS Biol.* **4**, e315 (2006).

9. Adams, R. H. *et al.* Roles of ephrinB ligands and EphB receptors in cardiovascular development: demarcation of arterial/venous domains, vascular morphogenesis, and sprouting angiogenesis. *Genes Dev.* **13**, 295–306 (1999).
10. Corada, M. *et al.* Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nature Commun.* **4**, 2609 (2013).
11. Adams, R. H. & Alitalo, K. Molecular regulation of angiogenesis and lymphangiogenesis. *Nature Rev. Mol. Cell Biol.* **8**, 464–478 (2007).
12. Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nature Med.* **6**, 389–395 (2000).
13. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
14. Hellström, M., Kalén, M., Lindahl, P., Abramsson, A. & Betsholtz, C. Role of PDGF-B and PDGFR- β in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126**, 3047–3055 (1999).
15. Chambers, S. M. *et al.* Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS Biol.* **5**, e201 (2007).
16. Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. *Nature Med.* **2**, 1011–1016 (1996).
17. Prisby, R. D. *et al.* Aging reduces skeletal blood flow, endothelium-dependent vasodilation, and NO bioavailability in rats. *J. Bone Miner. Res.* **22**, 1280–1288 (2007).
18. Moerman, E. J., Teng, K., Lipschitz, D. A. & Lecka-Czernik, B. Aging activates adipogenic and suppresses osteogenic programs in mesenchymal marrow stroma/stem cells: the role of PPAR- γ 2 transcription factor and TGF- β /BMP signaling pathways. *Aging Cell* **3**, 379–389 (2004).
19. Zsebo, K. M. *et al.* Stem cell factor is encoded at the Sl locus of the mouse and is the ligand for the c-kit tyrosine kinase receptor. *Cell* **63**, 213–224 (1990).
20. Vooijs, M. *et al.* Mapping the consequence of Notch1 proteolysis *in vivo* with NIP-CRE. *Development* **134**, 535–544 (2007).
21. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
22. Wu, G. *et al.* SEL-10 is an inhibitor of notch signaling that targets notch for ubiquitin-mediated protein degradation. *Mol. Cell. Biol.* **21**, 7403–7415 (2001).
23. Roca, C. & Adams, R. H. Regulation of vascular morphogenesis by Notch signaling. *Genes Dev.* **21**, 2511–2524 (2007).
24. Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nature Rev. Cancer* **3**, 721–732 (2003).
25. Red-Horse, K., Ueno, H., Weissman, I. L. & Krasnow, M. A. Coronary arteries form by developmental reprogramming of venous cells. *Nature* **464**, 549–553 (2010).
26. Xu, C. *et al.* Arteries are formed by vein-derived endothelial tip cells. *Nature Commun.* **5**, 5758 (2014).
27. Kinashi, T. & Springer, T. A. Steel factor and c-kit regulate cell-matrix adhesion. *Blood* **83**, 1033–1038 (1994).
28. Miyazawa, K. *et al.* Membrane-bound Steel factor induces more persistent tyrosine kinase activation and longer life span of c-kit gene-encoded protein than its soluble form. *Blood* **85**, 641–649 (1995).
29. Poulos, M. G. *et al.* Endothelial Jagged-1 is necessary for homeostatic and regenerative hematopoiesis. *Cell Reports* **4**, 1022–1034 (2013).
30. Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M. & de Haan, G. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J. Exp. Med.* **208**, 2691–2703 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Schiller for technical assistance, A. Starsichova for help with bone sample processing, M. Stehling for fluorescence-activated cell sorting (FACS), S. Volkery for microscopy, and M. Vanlandewijck and K. Nahar for help with *Pdgfr^{OE-EC}* bones. Funding was provided by the Max Planck Society, the University of Münster, the DFG cluster of excellence ‘Cells in Motion’, and the European Research Council (AdG 339409 ‘AngioBone’). This research was partly supported by a joint grant from the Ministry of Science, Technology & Space, Israel, DKFZ Germany, ERC AdG 294556 ‘BBBarrier’, the Knut and Alice Wallenberg Foundation and the Swedish Cancer Foundation.

Author Contributions A.P.K., S.K.R. and R.H.A. designed experiments and interpreted results. A.P.K. and S.K.R. organized and conducted most experiments, including generation and characterization of mouse lines, imaging, flow cytometric analysis and transplantations. T.I. and T.L. designed and performed transplantation experiments. M.A.M. and C.B. generated and provided samples from *Pdgfr^{OE-EC}* mice. A.P.K. and U.H.L. analysed *Efnb2* and *NICD-Cre* mice. A.P.K., S.K.R. and R.H.A. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H.A. (ralf.adams@mpi-muenster.mpg.de) or A.P.K. (anjali.kusumbe@mpi-muenster.mpg.de).

METHODS

Genetically modified and aged mice. C57BL/6J males were used for analysis of wild-type bone unless stated otherwise. Mice at the age of 2–5 weeks and 55–70 weeks were chosen for young and aged group sets, respectively. All EC-specific mutants were generated using *Cdh5(PAC)-CreERT2* transgenic mice unless indicated otherwise. For gene inactivation in the postnatal endothelium, mice carrying loxP-flanked *Rbpj* (*Rbpj^{lox/lox}*) alleles³¹ and *Cdh5(PAC)-CreERT2* transgenics³² were interbred. To induce Cre activity and gene inactivation, offspring were injected with 500 µg tamoxifen (Sigma, T5648) intraperitoneally every day from postnatal day (P)10 to P14. The resulting *Rbpj^{ΔEC}* (*CreERT2^{T/+}* *Rbpj^{lox/lox}*) mutants and Cre-negative littermate controls were killed at P28, and femurs and tibiae were collected for analysis. Identical breeding and tamoxifen administration strategies were used to generate EC-specific mutants with *Fbxw7^{lox/lox}* (ref. 33) or *Dll4^{lox/lox}* mice³⁴.

For EC-specific *Hif1a* deletions, *Cdh5(PAC)-CreERT2* transgenic mice were interbred with conditional *Hif1a* (*Hif1a^{lox/lox}*) mutants³⁵. To induce Cre activity and gene inactivation, pups were injected with 500 µg tamoxifen (Sigma, T5648) intraperitoneally every day from P10 to P14. Femurs and tibiae from *Cdh5(PAC)-CreERT2^{T/+}* *Hif1a^{lox/lox}* (*Hif1a^{ΔEC}*) mutants and Cre-negative *Hif1a^{lox/lox}* (Controls) were collected on P20 after euthanasia. The same approach was used for experiments involving conditional *Vhl* mice³⁶.

For *Fbxw7* deletion in the vasculature of aged mice, we generated litters with *Fbxw7^{lox/lox}* *Cdh5(PAC)-CreERT2^{T/+}* (*Fbxw7^{ΔEC}*) and *Fbxw7^{lox/lox}* (control) genotypes. To induce Cre activity and gene inactivation, 55- to 65-week-old mice were injected with 1,000 µg tamoxifen (Sigma, T5648) intraperitoneally every day for 5 days. After a rest period of 16 days, mice were subjected to a second round of tamoxifen injections with the same dosage and frequency as described above. After a further 16 days, mice were analysed after euthanasia.

For overexpression of the Notch1 intracellular domain (NICD), *Gt(ROSA)26Sor^{tm1(Notch1)Dam/J}* mice³⁷ and *Cdh5(PAC)-CreERT2* transgenics were interbred. Tamoxifen administration (see above for injection schedule) was used to generate *CreERT2*-positive (*NICD^{ΔEC}*) mutants overexpressing NICD in ECs and corresponding controls. For EC-specific PDGFB overexpression, *Rosa26-hPDGF-B* mice³⁸ were interbred with *Cdh5(PAC)-CreERT2* or with *Tie2 Cre³⁹* transgenics. To study the interplay between Notch and HIF signalling in ECs, endothelial specific double-mutant mice were generated using *Cdh5(PAC)-CreERT2* transgenics. *Cdh5(PAC)-CreERT2* mice were interbred with mice carrying the indicated combinations of *Hif1a^{lox/lox}*, *Rbpj^{lox/lox}*, *Vhl^{lox/lox}* and *Gt(ROSA)26Sor^{tm1(Notch1)Dam/J}* alleles.

For the detection of Notch cleavage and activity, *Notch1^{tm3(crr)Rko/J}* (*NICD-Cre*) mice²⁰, which carry a Cre recombinase fused to the carboxy (C) terminus of the intracellular domain of Notch1, were mated with *Rosa26-mG/mT* reporter animals²¹. The resulting *Notch1^{tm3(crr)Rko/J}* *R26-mG/mT^{T/+}* double heterozygotes were killed and analysed at 3 weeks of age. Genetic labelling of cells expressing ephrin-B2 was performed using B6;129S4-*Efnb2^{tm2Sor/J}* (*Efnb2^{GFP/+}*) knock-in mice, which express H2B-GFP under control of the endogenous *Efnb2* promoter⁸.

For labelling of proliferating cells, mice were intraperitoneally injected with 300 µg of EdU (Invitrogen) 3 h before euthanasia. Tibiae were immediately collected and processed. BM cells and bone sections were stained for EdU using Click-iT chemistry following the manufacturer's instructions (Invitrogen).

For metabolic labelling with the hypoxia probe pimonidazole (Pimo, Hypoxyprobe), mutant and control mice were intraperitoneally injected with 60 mg/kg body weight Pimo at 2 h before euthanasia. Metabolized Pimo was detected by a rabbit antiserum against the non-oxidized, protein-conjugated form of pimonidazole (Hypoxyprobe).

All animals were genotyped by PCR. Protocols and primer sequences can be provided upon request. Experiments involving animals were performed according to the institutional guidelines and laws, following protocols approved by local animal ethics committees.

Immunostaining of bone sections and cells. Freshly dissected bone tissues collected from wild-type mice or from mutants and their control littermates were immediately fixed in ice-cold 4% paraformaldehyde solution for 4 h. Decalcification was performed with 0.5 M EDTA at 4°C with constant shaking and decalcified bones were immersed into 20% sucrose and 2% polyvinylpyrrolidone (PVP) solution for 24 h. Finally, the tissues were embedded and frozen in 8% gelatin (porcine) in presence of 20% sucrose and 2% PVP. For immunofluorescent stainings and morphological analyses, sections were generated using low-profile blades on a Leica CM3050 cryostat.

For phenotypic analysis, mutant and littermate control samples were always processed, sectioned, stained, imaged and analysed together at the same conditions and settings. For immunostaining, bone sections were air-dried, permeabilized for 10 min in 0.3% Triton X-100, blocked in 5% donkey serum at room temperature

for 30 min and probed with the primary antibodies diluted in 5% donkey serum in PBS for 2 h at room temperature or overnight at 4°C.

After primary antibody incubation (Supplementary Table 1), sections were washed with PBS three times and incubated with appropriate Alexa Fluor-coupled secondary antibodies (1:400, Molecular Probes) for 1 h at room temperature. Nuclei were counterstained with DAPI. Sections were thoroughly washed with PBS before mounting them using FluoroMount-G (Southern Biotech). Finally, cover slips were sealed with nail polish.

Immunostaining of sorted HSCs and progenitor cells was performed as described previously⁴⁰. Briefly, cells were pipetted onto poly-lysine coated slides, incubated for 10 min, fixed with 4% PFA for 10 min at room temperature, permeabilized in 0.15% Triton X-100 for 2 min at room temperature and blocked in 2% donkey serum overnight at 4°C. Slides were then incubated for 2 h with the anti-phospho-H2AX, washed thrice and incubated with the appropriate secondary antibody.

Image acquisition and quantitative analysis. Immunofluorescent stainings were analysed at high resolution with a Zeiss laser scanning confocal microscope, LSM-780. Z-stacks of images were processed and reconstructed in three dimensions with Imaris software (version 7.00, Bitplane). Imaris, Photoshop and Illustrator (Adobe) software were used for image processing in compliance with Nature's guidelines for digital images. All quantifications were done with ImageJ and Imaris software on high-resolution confocal images.

Quantitative reverse transcription PCR. For the analysis of messenger RNA (mRNA) expression levels in type H or type L endothelium, CD31^{hi} Emcn^{hi} and CD31^{lo} Emcn^{lo} cells were sorted by FACS directly into the lysis buffer of the RNeasy Mini Kit (QIAGEN). Total RNA was isolated according to the manufacturer's protocol. A total of 100 ng RNA per reaction was used to generate complementary DNA (cDNA) with the iScript cDNA Synthesis System (Bio-Rad). qPCR was performed using TaqMan gene expression assays on an ABI PRISM 7900HT Sequence Detection System. The FAM-conjugated TaqMan probes *Efnb2* and *Sox17* were used along with TaqMan Gene Expression Master Mix (Applied Biosystems). Gene expression assays were normalized to endogenous VIC-conjugated *Actb* probes as standard. For analysis of mRNA expression levels from whole bones, dissected femurs or dissected metaphysis (as described in the figure legends) were immediately crushed finely, digested with collagenase and centrifuged to obtain a pellet, which was then lysed into lysis buffer of RNeasy Mini Kit (QIAGEN). For cells in culture, culture medium was completely removed and cells were immediately lysed with lysis buffer. A total of 500 ng RNA per reaction was used to generate cDNA with the iScript cDNA Synthesis System (Bio-Rad), which was further processed as described above. FAM-conjugated TaqMan probes *Sp7*, *Cspg4*, *Pdgfrb*, *Sp7*, *Acan*, *Cfd*, *Hif1a*, *Epas1*, *Cxcl12*, *Fgf1*, *Kitl*, *Tgfb3*, *Tgfb1* and *Vegfa* were used along with TaqMan Gene Expression Master Mix (Applied Biosystems) to perform qPCR.

Flow cytometry. For flow cytometric analysis and sorting of type H and type L ECs, tibiae and femurs were collected, cleaned thoroughly to remove the adherent muscles. The epiphysis was removed and only the metaphysis and diaphysis regions were processed. Tibiae were then crushed in ice cold PBS with mortar and pestle. Whole BM was digested with collagenase incubation at 37°C for 20 min. Equal number of cells were then subjected to immunostaining with Emcn antibody (Santa Cruz, sc-65495) for 45 min. After washing, cells were stained with APC-conjugated CD31 antibody (R&D Systems, FAB3628A) for 45 min and phycoerythrin-conjugated secondary anti-rat antibody. After washing, cells were acquired on a BD FACS Canto flow cytometer or BD FACSVerser and analysed using BD FACSDiva (version 6.0, BD Bioscience) or BD FACSuite software. Cell sorting was performed with a BD FACS Aria II.

For demarcating and sorting CD31^{hi} Emcn^{hi} ECs, first standard quadrant gates were set. Subsequently, to differentiate CD31^{hi} Emcn^{hi} cells from the total double positive cells in quadrant 2, gates were arbitrarily set at >10⁴ log(FI-4 (CD31-APC) fluorescence) and >10⁴ log(FI-2 (endomucin-PE) fluorescence).

For the analysis of total ECs in bone, tibiae were processed as described above to obtain single-cell suspensions, which were stained with biotin-coupled CD45 (BD, 553077) or Ter119 (BD, 559971) antibodies for 45 min. After washing in PBS, cells were stained with Streptavidin PE-Cy5 (BD, 554062) and Alexa Fluor488-conjugated CD31 (R&D Systems, FAB3628G) antibodies for 45 min. After washing, cells were acquired on FACS Canto and FACS Verse flow cytometers and analysed using FACSDiva (version 6.0, BD Bioscience) and FACS Suite software respectively. Total bone ECs were quantified as CD31⁺/CD45⁺/Ter119⁺. Endomucin was used to distinguish Emcn⁺ arterial ECs from Emcn⁺ sinusoidal and venous cells.

For the analysis of HSC frequency in the BM, BM cells were isolated by crushing the long bones with mortar and pestle in Ca²⁺- and Mg²⁺-free PBS 2% heat-inactivated bovine serum. The cells were drawn by passing through a 25-gauge needle several times and filtered with a 70-µm filter. The following antibodies were

used to stain HSCs: biotin-labelled lineage markers (CD5, CD11b, CD45R, Gr-1 and Ter119), cKit, Sca-1, CD48 and CD150 antibodies (Supplementary Table 1).

For the enrichment and sorting of HSCs and progenitor cells, BM cells were isolated by crushing the long bones with mortar and pestle in Ca^{2+} - and Mg^{2+} -free PBS 2% heat-inactivated bovine serum. The cells were drawn by passing through a 25-gauge needle several times and filtered with a 70- μm filter. The obtained single-cell suspension obtained was subjected to lineage depletion (MACS, Miltenyi Biotech). Lineage-depleted BM cells were then stained with cKit and Sca1 antibodies (Supplementary Table 1). After washing, cell sorting was performed with a BD FACS Aria II.

For the analysis of perivascular cells, Sca1⁺ ECs, ephrin-B2⁺ ECs and HSC frequency, the above-described protocol was used to obtain a single-cell suspension followed by immunostaining with the appropriate antibodies (Supplementary Table 1).

BM transplantation experiments. Competitive repopulation assays were performed using the CD45.1/CD45.2 congenic system. Equivalent volumes of BM cells collected from EC-specific mutant mice or littermate control mice (CD45.2) were transplanted into lethally irradiated (12 Gy) CD45.1 recipients with 0.3×10^6 competitor CD45.1 cells. CD45.1/CD45.2 chimaerism of recipient blood was analysed up to 7 months after transplantation using flow cytometry analysis. For the secondary transplantation (performed for *Fbxw7* ΔEC mutant mice and littermate controls), 1×10^6 BM cells from CD45.1 mice that had previously undergone transplantation at 1:1 ratio were isolated at 7 months after transplantation and injected into lethally irradiated recipients. For calculation of competitive repopulating units (CRU), recipient mice were transplanted with limiting dilutions of donor-derived BM cells (2.5×10^4 to 2×10^5) together with 2×10^5 recipient-derived BM cells. Mice were killed after 18 weeks and the multi-lineage myelo-lymphoid donor-derived contribution in the peripheral blood was assessed using flow cytometry analysis. HSC-CRU frequency and statistical significance was determined using ELDA software (<http://bioinf.wehi.edu.au/software/elda/>)^{41,42}.

Primary culture of ECs and PDGFR β ⁺ cells from bone. Tibiae and femurs from wild-type mice were collected in sterile Ca^{2+} - and Mg^{2+} -free PBS, crushed with mortar and pestle, subjected to collagenase digestion, and filtered and washed thrice to obtain a single-cell suspension. Endothelial cells were then sorted using endomucin antibody (catalogue number SC-65495) and Dynabeads sheep anti-Rat IgG (Invitrogen). Sorted ECs were then plated on dishes coated with fibronectin and cultured in endothelial cell growth medium (EBM-2, Clonetics; Lonza) supplemented with EGM-2 SingleQuots (CC-4176, Clonetic; Lonza).

PDGFR β ⁺ cells were sorted from single-cell suspensions using CD140b/PDGFR β Receptor β antibody (eBioscience, catalogue number 14-1402-82) and Dynabeads sheep anti-Rat IgG (Invitrogen). Sorted PDGFR β ⁺ cells were cultured on tissue culture plates containing in alpha MEM (Gibco) and 10% fetal bovine serum (Gibco).

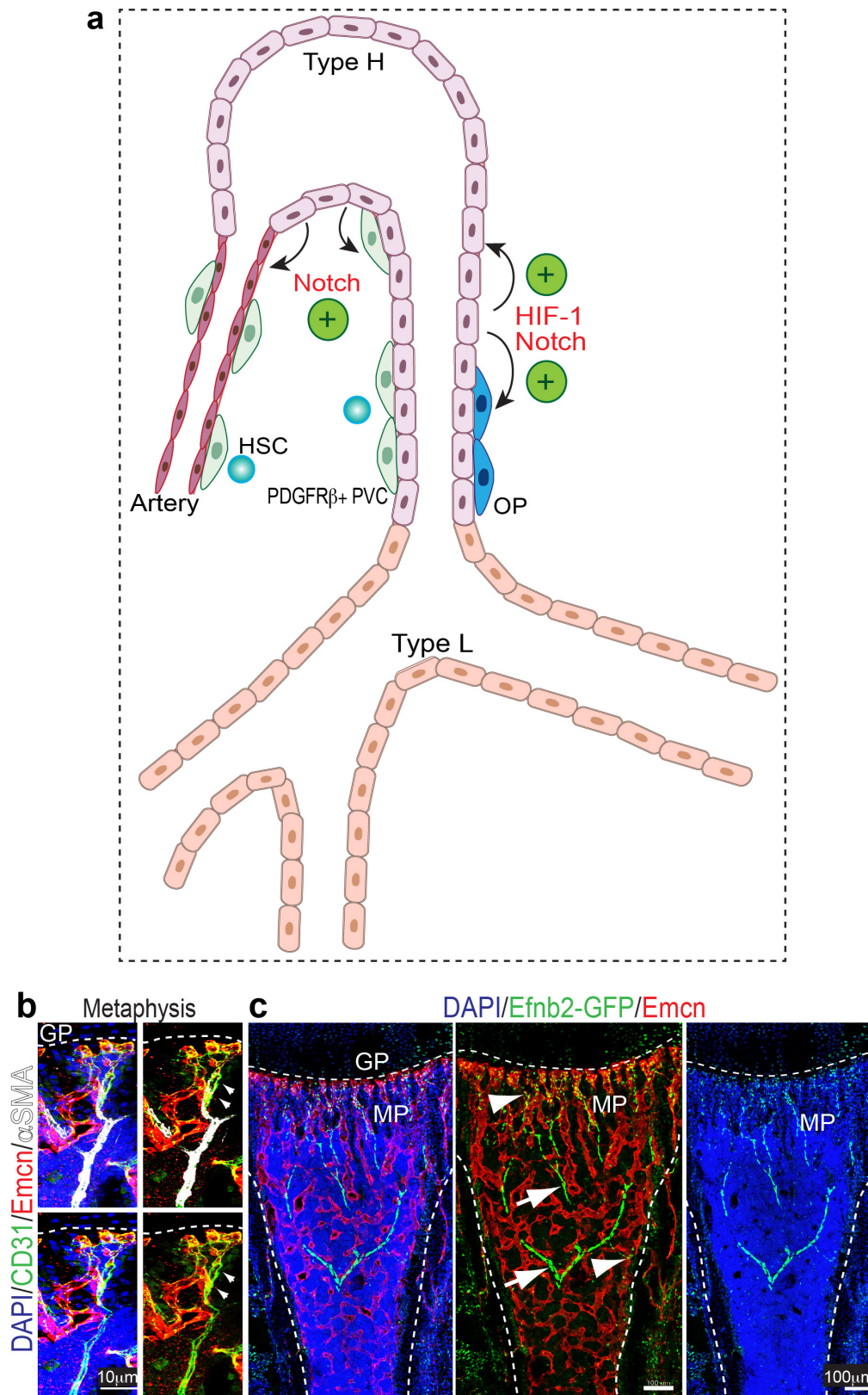
Cultures of ECs or PDGFR β ⁺ cells were maintained at 37 °C with 5% CO₂ in a humidified atmosphere. For DFM treatment and subsequent analysis, cultures between passage 1 and 2 were used. Cells were treated with DFM (6.25 mg ml⁻¹ of culture medium) for the duration of 36 h and subsequently cell culture medium (directly) or cells after trypsinization, three washings and lysate preparation were used for ELISA.

ELISA. SCF levels in the mice bone supernatant/extracellular fluid (secreted SCF) or cell culture medium (secreted/extracellular SCF) or cell lysate from cultured cells (cellular/membrane-bound SCF) or cell lysate (cellular/membrane-bound SCF) prepared from the single-cell suspension of femurs (see primary culture section) were determined by ELISA kits (Sigma-Aldrich and USCN Business).

CFU-F and MSC differentiation assays. CFU-F assay was performed as described previously¹⁶. MSC differentiation into osteogenic, chondrogenic and adipogenic lineages was performed using a MSC functional identification kit (R&D Systems) according to the manufacturer's instructions. Fourteen days after differentiation, we performed qPCR analysis.

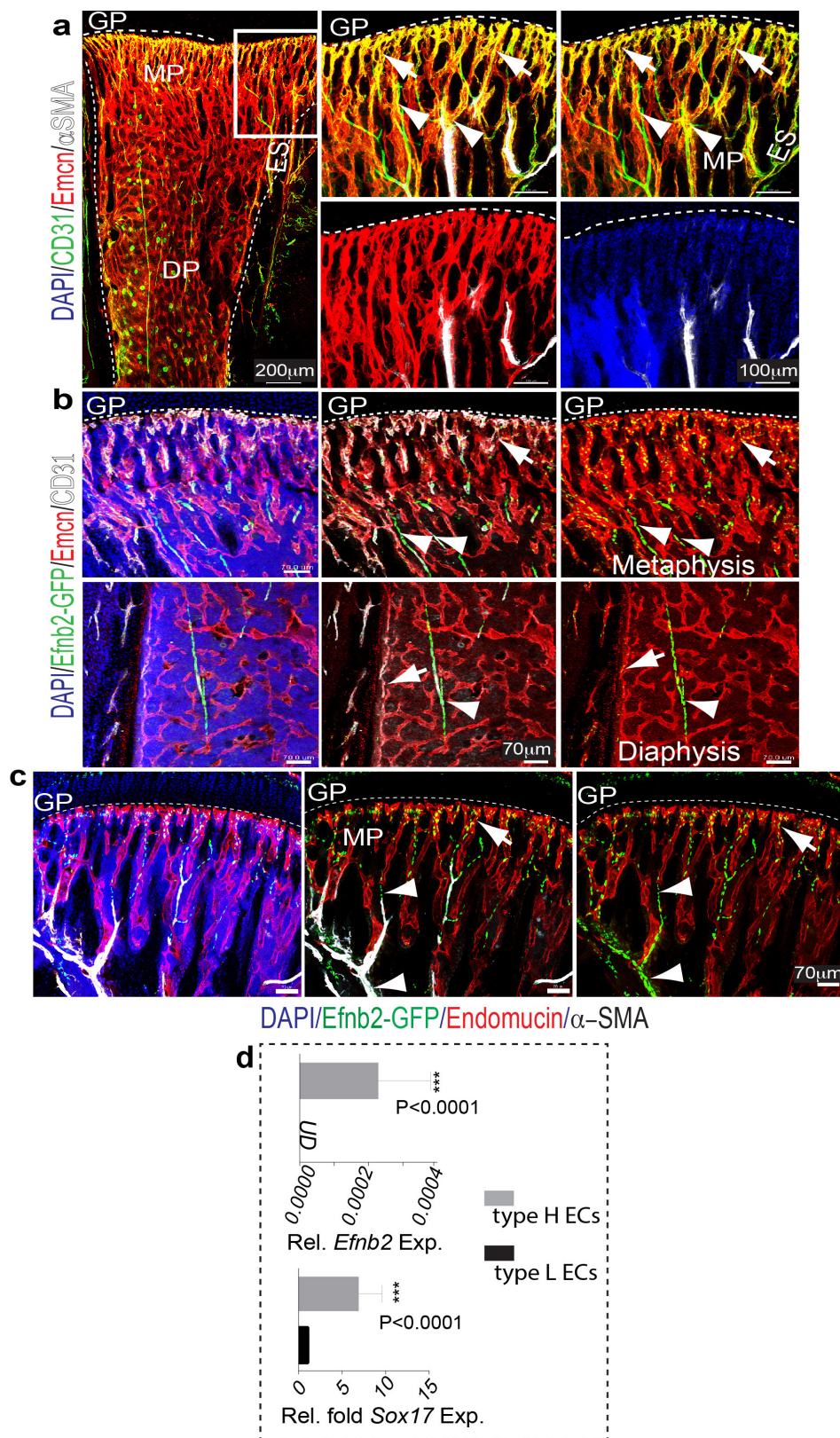
Statistical analysis. All data are presented as either mean \pm s.e.m. or mean \pm s.d. (as indicated in figure legends). The data presented in the figures reflect multiple independent experiments performed on different days using different mice. Unless otherwise mentioned, most of the data presented in figure panels are based on three independent experiments. The significance of difference was determined using a two-tailed Student's *t*-test unless otherwise mentioned. $P > 0.05$ was considered not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Student's *t*-test with Welch's correction was performed when group sizes were not equal. For analysis of the statistical significance of differences between more than two groups, we performed repeated measures one-way analysis of variance (ANOVA) tests with Greenhouse-Geisser correction (variances between groups were not equal) and Tukey's multiple comparison tests to assess statistical significance with a 95% confidence interval. In all the figures, *n* refers to the number of mice. For data with unequal group sizes, the first numerical value for *n* refers to mutants and the second refers to control mice. All statistical analyses were performed using Graphpad Prism software. No randomization or blinding was used and no animals were excluded from analysis. Sample sizes were selected on the basis of previous experiments. Unless otherwise indicated, results are based on three independent experiments to guarantee reproducibility of findings.

- Han, H. *et al.* Inducible gene knockout of transcription factor recombination signal binding protein-J reveals its essential role in T versus B lineage decision. *Int. Immunol.* **14**, 637–645 (2002).
- Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
- Hoek, J. D. *et al.* Fbw7 controls neural stem cell differentiation and progenitor apoptosis via Notch and c-Jun. *Nature Neurosci.* **13**, 1365–1372 (2010).
- Koch, U. *et al.* Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J. Exp. Med.* **205**, 2515–2523 (2008).
- Tang, N. *et al.* Loss of HIF-1 α in endothelial cells disrupts a hypoxia-driven VEGF autocrine loop necessary for tumorigenesis. *Cancer Cell* **6**, 485–495 (2004).
- Haase, V. H., Glickman, J. N., Socolovsky, M. & Jaenisch, R. Vascular tumors in livers with targeted inactivation of the von Hippel-Lindau tumor suppressor. *Proc. Natl Acad. Sci. USA* **98**, 1583–1588 (2001).
- Murtaugh, L. C., Stanger, B. Z., Kwan, K. M. & Melton, D. A. Notch signaling controls multiple steps of pancreatic differentiation. *Proc. Natl Acad. Sci. USA* **100**, 14920–14925 (2003).
- Armulik, A. *et al.* Pericytes regulate the blood-brain barrier. *Nature* **468**, 557–561 (2010).
- Kisanuki, Y. Y. *et al.* Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis *in vivo*. *Dev. Biol.* **230**, 230–242 (2001).
- Itkin, T. *et al.* FGF-2 expands murine hematopoietic stem and progenitor cells via proliferation of stromal cells, c-Kit activation, and CXCL12 down-regulation. *Blood* **120**, 1843–1855 (2012).
- Miller, C. L., Dykstra, B. & Eaves, C. J. Characterization of mouse hematopoietic stem and progenitor cells. *Curr. Protoc. Immunol.* **22**, 22B.2.1–22B.2.31 (2008).
- Gomez, C. R., Knutson, G. J., Clifton, K. B., Schreiber, C. A. & Vuk-Pavlović, S. Age-dependent response of murine female bone marrow cells to hyperbaric oxygen. *Biogerontology* **13**, 287–297 (2012).



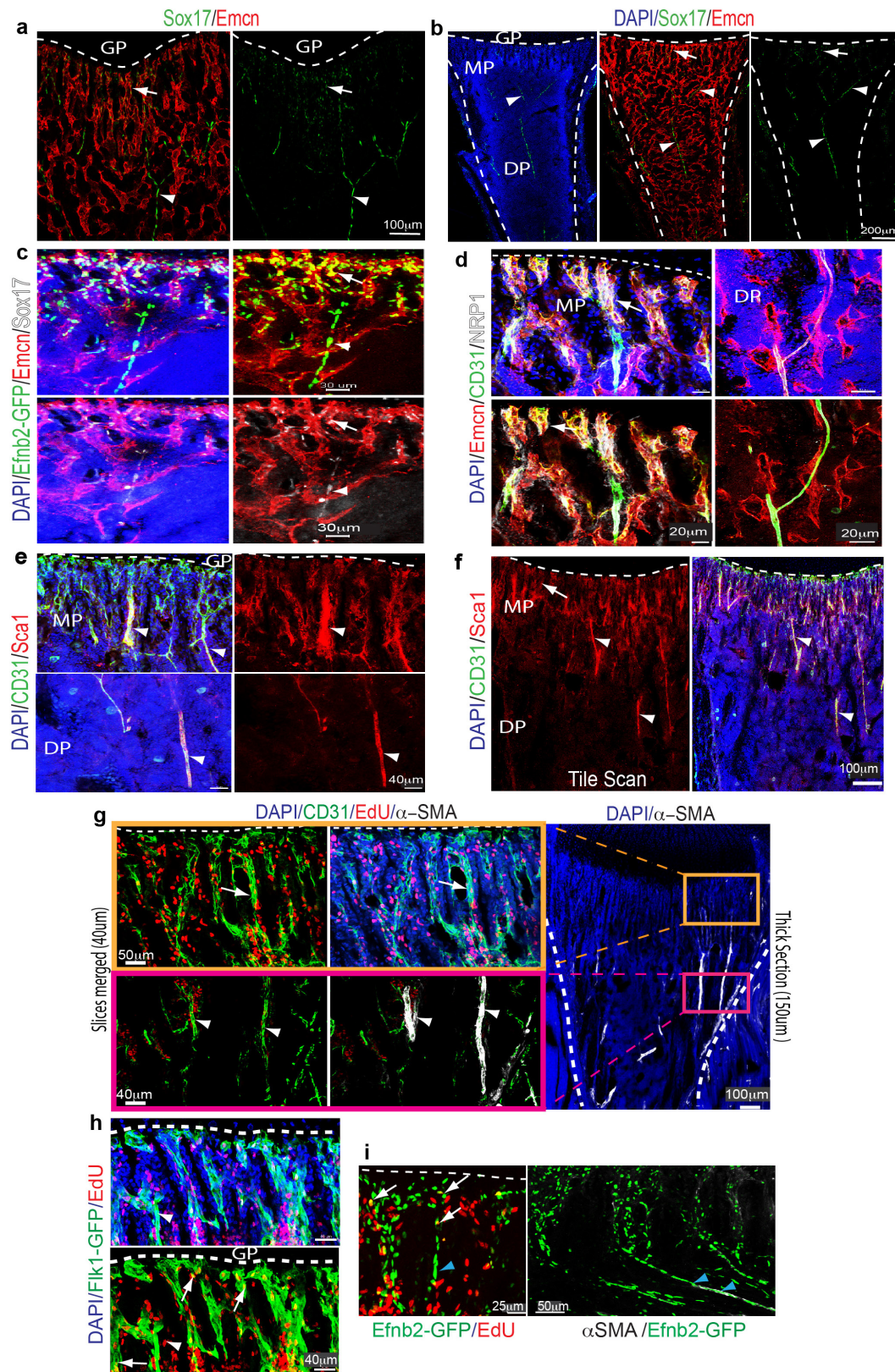
Extended Data Figure 1 | Summary of key findings and analysis of vessel subtypes. **a**, Type H vessels express arterial markers and, unlike type L (sinusoidal) vessels, make direct connections to distal arterioles. PDGFR β ⁺ perivascular cells (PVC) are selectively abundant around type H vessels and endosteal arteries. In addition, HSCs are frequently detected near the type H endothelium and arterioles in the endosteal region. Endothelial Notch signalling drives artery formation, expansion of type H ECs and PDGFR β ⁺ PVCs, production of cell-bound SCF and thereby enhances vascular HSC niche function. By contrast, endothelial HIF

signalling fails to induce artery formation and expansion of PDGFR β ⁺ PVCs, which does not result in more HSC niches despite of increases in type H ECs and Osterix⁺ osteoprogenitor (OP) cells. **b**, Confocal images of 8-week-old tibial metaphysis showing the direct connection between a CD31^{hi} (green) Emcn⁺ (red) α -SMA⁺ artery (white) and CD31^{hi} Emcn^{hi} type H capillaries. Dashed line, growth plate. **c**, Representative confocal tile scan of 2-week-old *Efnb2*^{GFP/+} (green) tibia stained with Emcn (red). Note GFP signal in arteries (arrows) and type H vessels in metaphysis and endosteum (arrowhead), while type L vessels lack *Efnb2*^{GFP/+} expression.



Extended Data Figure 2 | Relationship between type H capillaries and arterioles. **a**, Tile scan (left) and different representations (merged channels) of image data in inset. Shown is a 4-week-old tibial metaphysis immunostained for CD31 (green), Emcn (red) and α -SMA (white). CD31^{hi} (green) Emcn⁻ arteries (arrowheads) are directly connecting to type H ECs (arrows). ES, endosteum. **b**, Maximum intensity projections of 4-week-old *Efnb2*^{GFP/+} (GFP, green) tibial metaphysis and diaphysis immunostained for CD31 (white) and Emcn (red). Panels in centre and on the right different representations (merged channels) of image data. Only

Emcn⁻ arterial ECs (arrowheads) and Emcn^{hi} type H ECs in metaphysis and endosteum (arrows) are positive for GFP, while diaphyseal type L ECs lack *Efnb2*^{GFP} expression. **c**, Confocal images of *Efnb2*^{GFP/+} metaphysis showing GFP⁺ (green) Emcn^{hi} (red) type H ECs (arrows) connecting to α -SMA⁺ (white) cell-covered, Emcn⁻ artery (arrowhead). Nuclei, DAPI (blue). **d**, RT-qPCR analysis of *Efnb2* and *Sox17* expression (normalized to *Actb*) in sorted tibial CD31^{hi} Emcn^{hi} relative to CD31^{lo} Emcn^{lo} ECs. Data represent mean \pm s.d. ($n = 6$; three independent experiments). P values, two-tailed unpaired t -test. UD, undetectable.

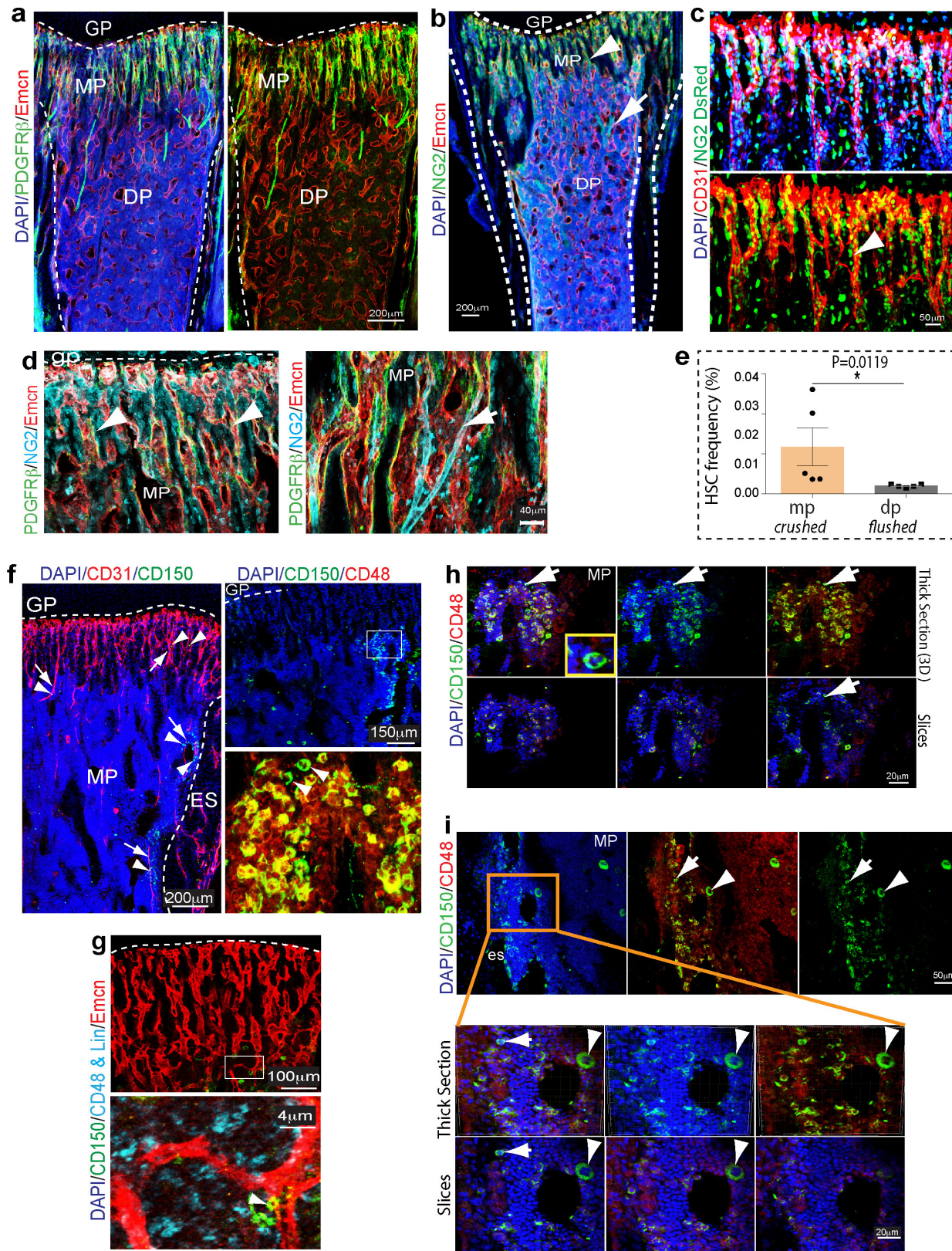


Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Characterization of bone vessel populations.

a, b, Confocal (**a**) and tile scan confocal (**b**) images showing *Emcn* (red) and *Sox17* (green) expression in 4-week-old tibia. Dashed line marks growth plate in the metaphysis. *Sox17* is expressed by type H ECs (arrows) and arteries (arrowheads), whereas signal is absent in type L vessels of the diaphysis. **c**, Confocal images showing GFP (green), *Emcn* (red) and *Sox17* (white) expression in the metaphysis region and diaphysis region of a *Efnb2*^{GFP/+} tibial section. Nuclei, DAPI (blue). Note *Sox17* and GFP double-positive cells in type H capillaries (arrow) connecting to *Efnb2*^{GFP+} *Emcn*⁻ artery (arrowhead). **d**, Representative confocal images from metaphysis (left) and diaphysis (right) regions of 4-week-old murine long bone showing *Emcn* (red), CD31 (green) and NRP1 (white) immunostaining. Nuclei, DAPI (blue). NRP1 marks type H capillaries (arrow) and arteries, while type L vessels lack detectable NRP1 expression. **e, f**, Representative confocal (**e**) and tile scan confocal (**f**) images showing Sca1 (red) and CD31 (green) immunostaining in metaphysis and diaphysis

of 4-week-old tibia. Sca1 staining decorates CD31^{hi} arteries (arrowheads) and type H ECs (arrows) but not type L ECs. Dashed line marks border of the growth plate. **g**, Maximum intensity projections of EdU (red) labelled 3-week-old tibia immunostained for CD31 (green) and α -SMA (white). Nuclei, DAPI (blue). Proliferating (EdU⁺) CD31^{hi} type H ECs (arrow) were found near the growth plate (dashed line), whereas distal α -SMA⁺ cell-covered arteries (arrowheads) lack EC proliferation. **h**, Confocal images showing EdU labelling in 5-week-old *Flk1-GFP* long bone. Note EdU⁺ and GFP⁺ (arrows) in endothelial columns close to the growth plate. Straight, small calibre arterioles (arrowheads) lack proliferating ECs. **i**, Confocal images of 3-week-old *Efnb2*^{GFP/+} (green) tibia sections with α -SMA immunostaining (white). Ephrin-B2⁺ arteriolar ECs exhibit elongated nuclei (blue arrowheads), which are not typical for type H ECs. Overlap of GFP and EdU signals in type H capillaries (arrows) but not in arteries (blue arrowheads).

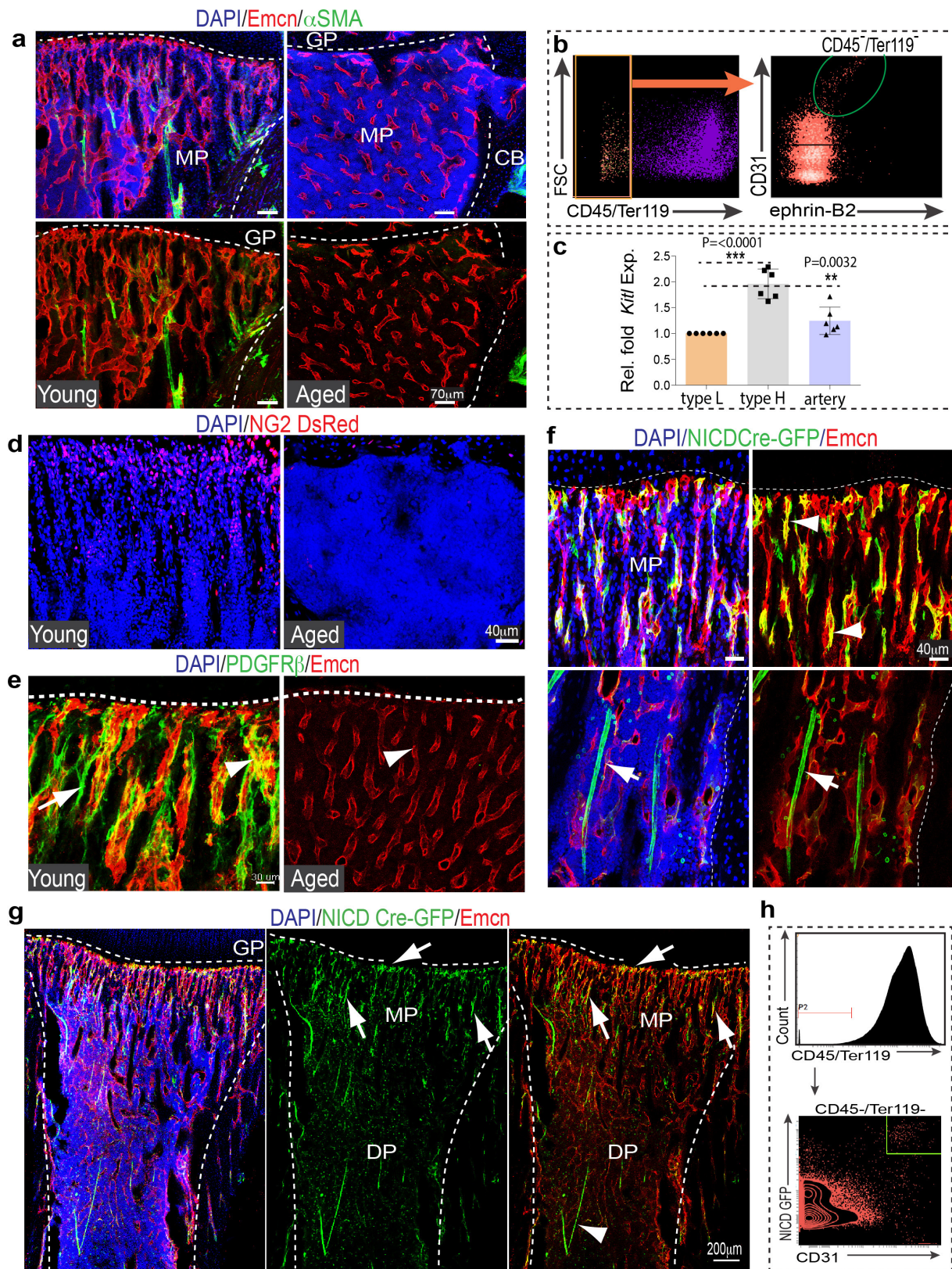


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Distribution of perivascular cells and HSCs.

a, PDGFR β (green) cell localization in 4-week-old tibia. PDGFR β ⁺ cells surround arteries and Emcn^{hi} (red) metaphyseal type H vessels but not diaphyseal sinusoidal vessels. **b**, Representative tile scan images of 4-week-old long bone immunostained for Emcn (red) and NG2 (green). Nuclei, DAPI (blue). Dashed lines mark growth plate and compact bone. Note abundance of NG2 signals around type H vessel columns (arrowhead) in metaphysis and arteries (arrow). **c**, Confocal images of 4-week-old *NG2-DsRed* (DsRed shown as green) transgenic long bone after CD31 immunostaining (red). Nuclei, DAPI (blue). Note abundance of DsRed⁺ cells around CD31^{hi} (type H) capillaries (arrowhead). **d**, Sections of 3-week-old tibia immunostained for Emcn (red), NG2 (light blue) and PDGFR β (green), as indicated. PDGFR β ⁺ cells were abundant around arteries (arrow) and type H vessels (arrowheads) in metaphysis. **e**, Graph showing CD150⁺ CD48⁻ Lin⁻ Sca1⁺ c-Kit⁺ HSC frequency in single-cell suspension. The metaphysis region from long bones was dissected and, subsequently, these bone fragments were crushed and subjected to collagenase digestion. Simultaneously, CD150⁺ CD48⁻ Lin⁻ Sca1⁺ c-Kit⁺ HSC frequency was also quantitated for BM flushed from diaphysis. Data represent mean \pm s.e.m. ($n = 5$; three independent experiments). *P* values, two-tailed unpaired *t*-test. **f**, Maximum intensity projections showing CD31⁺ (red) arteries and type H vessels. CD150⁺ cells (green, arrowheads) in proximity of metaphyseal

and endosteal endothelium as well as CD31⁺ arterioles (arrowheads). Top right, distribution of CD150⁺ and CD48⁺ cells in 4-week-old tibial metaphysis. Arrowheads in higher magnification of inset (bottom) mark CD150⁺ and CD48⁻ HSCs. Dashed line marks growth plate. **g**, Perivascular localization of CD150⁺ (green) CD48⁻ Lineage⁻ (blue) HSCs (arrowhead) near Emcn⁺ (red) endothelium. **h**, Representative confocal image showing CD150⁺ cells (green) in the 4-week-old tibial metaphysis. Arrow marks rare CD150⁺ CD48⁻ cell, which is also shown at high magnification in inset. Top panels show three-dimensional reconstruction of a thick section, while optical slices (mimicking thin sections) are shown at the bottom. Note that CD150⁺ CD48⁻ cell (arrow) is only captured in one optical section and would therefore appear more isolated in thin tissue sections. **i**, Maximum intensity projections (top) showing CD150⁺ (green) cells in 4-week-old tibial metaphysis. Arrow marks rare CD150⁺ CD48⁻ cell, arrowhead indicates nearby CD150⁺ megakaryocyte. Higher magnifications of inset show three-dimensional projection and thin optical slices, as indicated. General abundance of CD150⁺ cells appears strongly reduced in individual optical sections and the indicated CD150⁺ CD48⁻ cell (arrow) is only captured in one slice. Notably, CD150⁺ cells were frequently found in clusters in thick cryosections (100 μ m), but appeared scattered in thin optical slices, which reflects the reported widespread distribution throughout the BM.

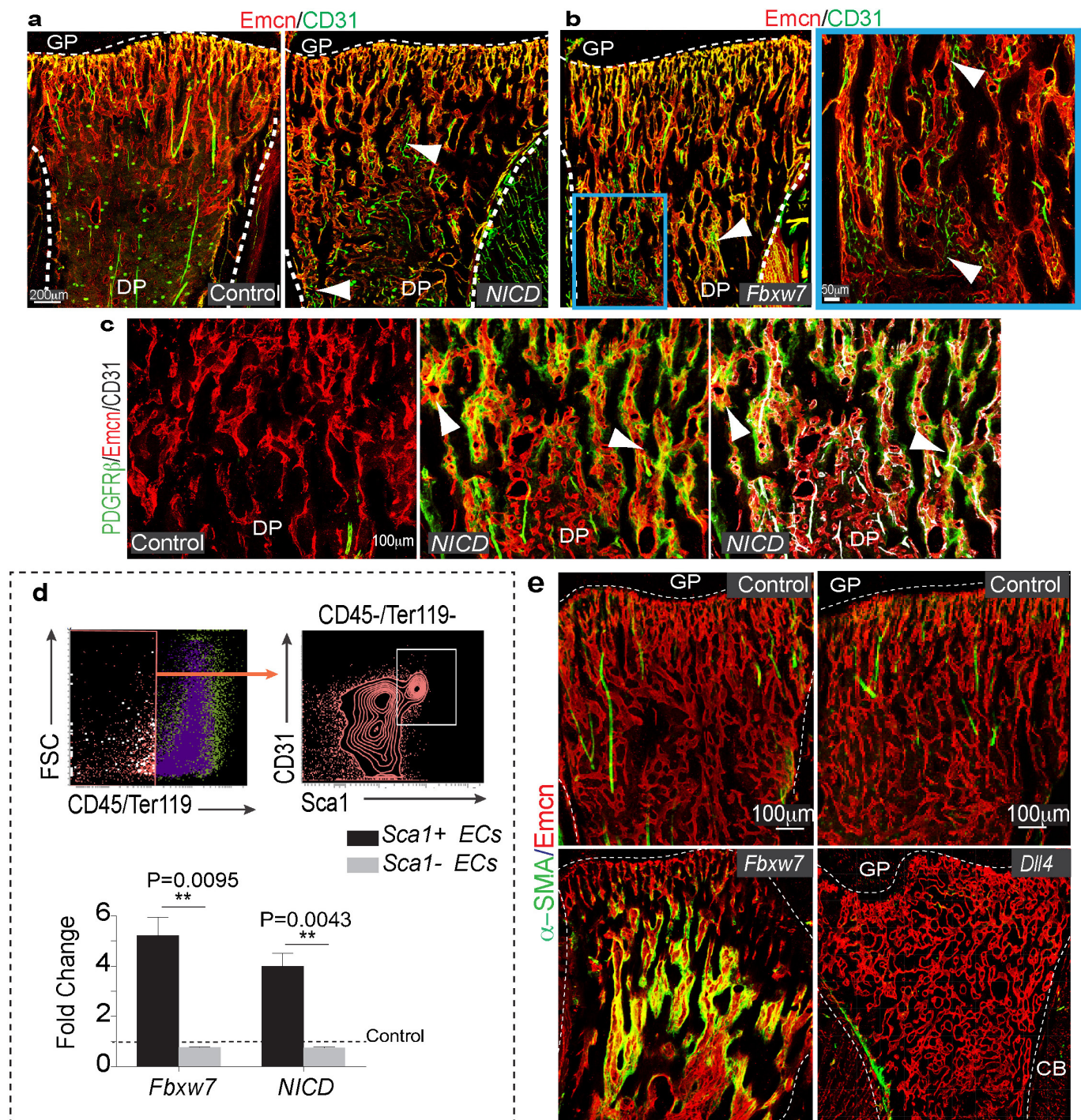


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Age-associated changes in bone vasculature.

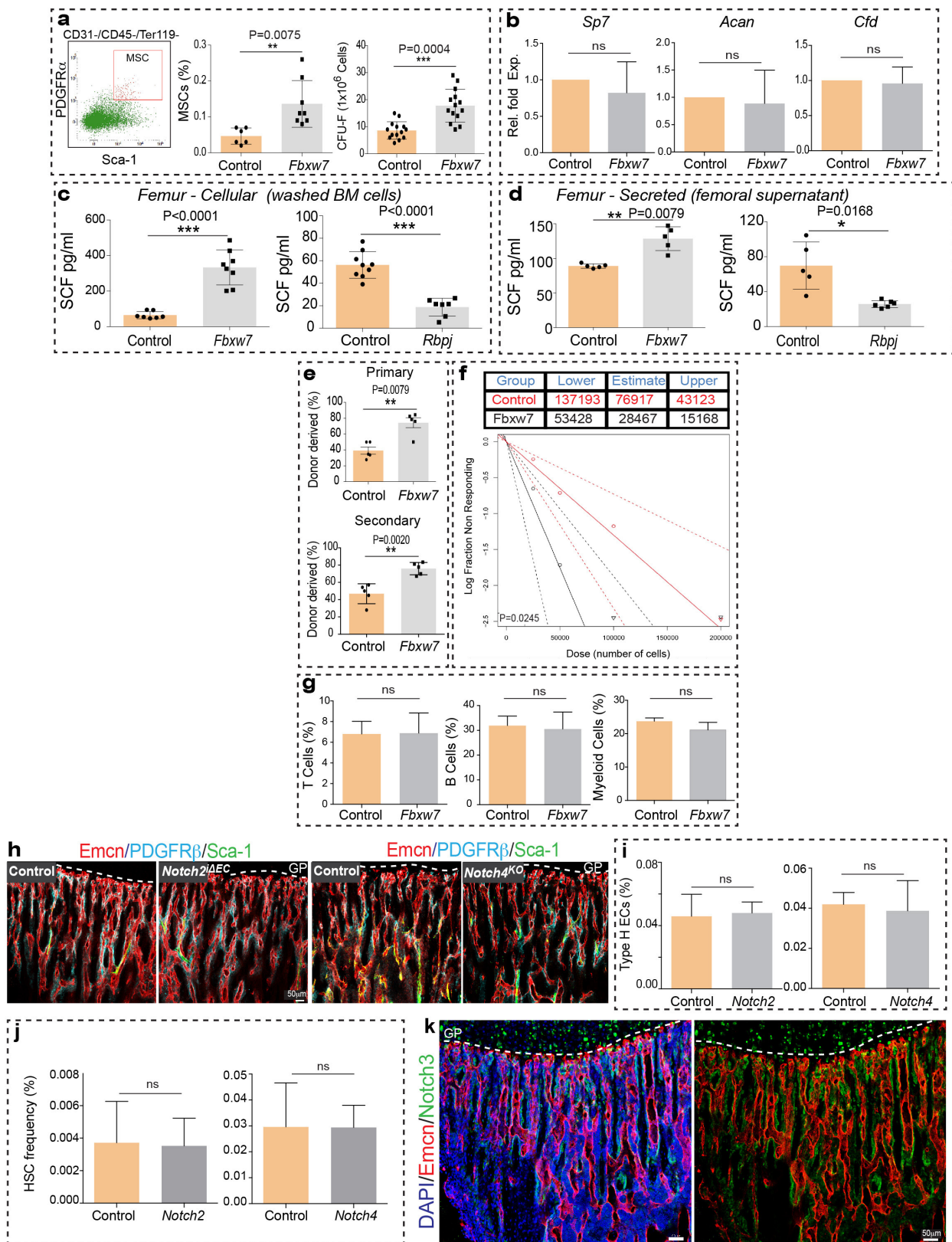
a, Representative confocal images from the metaphysis of young (3-week-old) long bone and the corresponding region in aged (70-week-old) long bone after α -SMA (green) and Emcn (red) immunostaining. Note decline in α -SMA⁺ cell-covered arteries in aged samples. Growth plate and compact bone are indicated. **b**, Dot plots showing ephrin-B2⁺ ECs sorted by flow cytometry from a single-cell suspension isolated from long bone. Total ECs in the single-cell suspension were identified as CD45⁻/Ter119⁻/CD31⁺ cells. **c**, qPCR analysis of *Kitl* expression (normalized to *Actb*) by CD31^{hi} Emcn^{hi} type H ECs and arterial ECs relative to CD31^{lo} Emcn^{lo} ECs sorted from murine tibia. Arterial ECs were identified and sorted as Emcn⁻ CD31⁺ ephrin-B2⁺ cells. Data represent mean \pm s.d. ($n = 6$; in three independent experiments). *P* values, two-tailed unpaired *t*-test. Note significantly higher *Kitl* expression in type H and arterial ECs relative to type L ECs. **d**, Confocal images showing young (4-week-old) and aged (65-week-old) *NG2-DsRed* (red) long bone. Nuclei, DAPI (blue). **e**, Maximum intensity projections of sectioned metaphyseal regions from

young (3-week-old) and aged (70-week-old) long bone immunostained for PDGFR β (green) and Emcn (red). Nuclei, DAPI (blue). Dashed lines indicated the border to growth plate. Arrowheads indicate capillaries and arrow shows an artery. **f**, Representative confocal images showing *NICD-Cre*-induced GFP expression (green) in Emcn⁺ (red) type H capillaries (arrowheads) and Emcn⁻ arteries (arrows) in 3-week-old tibia. Dashed lines mark growth plate or compact bone. **g**, Representative tile scan confocal images of 3-week-old tibia sections from *NICD-Cre* knock-in transgenic mice in the *Rosa26-mT/mG* reporter background. Vessels have been visualized by Emcn (red) immunostaining. Nuclei, DAPI (blue). Note high GFP expression in type H endothelium (arrows) and arteries (arrowheads). Dashed lines mark growth plate and endosteum; metaphysis and diaphysis are indicated. **h**, Contour plot showing intensities of CD31 immunostaining and GFP in single-cell suspension obtained from 3-week-old *NICD-Cre Rosa26-mT/mG* mice. ECs were demarcated as CD45⁻ Ter119⁻ CD31⁺. Note high GFP intensity in the CD31^{hi} EC subset.



Extended Data Figure 6 | Effect of endothelial Notch on the BM stroma. **a, b**, Confocal images showing CD31 (green) and Emcn (red) immunostained tibia sections of *NICD*^{ΔEC} (*NICD*) mutant and littermate control (**a**) or *Fbxw7*^{ΔEC} mutant (**b**). Small, interconnected arterioles (arrowheads) were abundant in Notch gain-of-function mutants. Growth plate, metaphysis and diaphysis are indicated. **c**, Representative confocal images from *NICD*^{ΔEC} mutant (*NICD*) and littermate control long bone after Emcn (red), CD31 (white) and PDGFRβ (green) immunostaining. Note strong accumulation of PDGFRβ⁺ cells in *NICD*^{ΔEC} bone (arrowheads). **d**, Quantitative analysis of Sca1⁺ ECs and Sca1⁻ ECs in

Fbxw7^{ΔEC} or *NICD*^{ΔEC} long bone relative to Cre-negative littermate controls. Data represent mean ± s.e.m. (*n* = 6). *P* values, two-tailed unpaired *t*-test. Total ECs were identified as CD45⁺ TER119⁻ CD31⁺ cells, as represented in the FACS plots. Note significant increase of Sca1⁺ ECs in mutants, whereas Sca1⁻ ECs remain comparable to control. **e**, Confocal tile scans of 4-week-old *Fbxw7*^{ΔEC}, *Dll4*^{ΔEC} and littermate control tibias immunostained with Emcn (red) and α-SMA (green). Unless otherwise mentioned, data presented in the panels are based on three independent experiments.

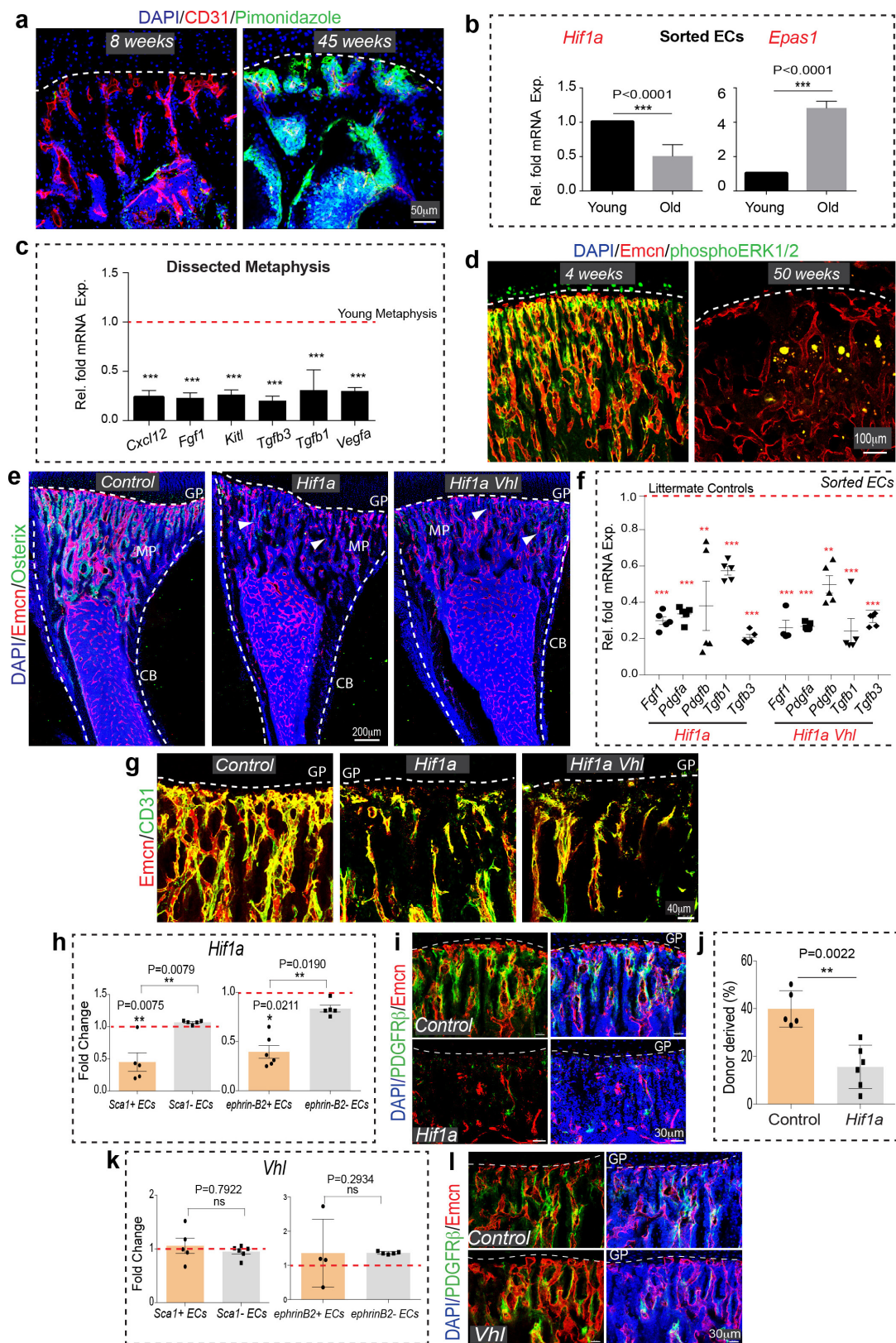


Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Endothelial Notch improves HSC niche function.

a, Representative dot plot showing the gating strategy used for defining the CD31⁺ CD45⁺ Ter119⁺ Sca-1⁺ and PDGFR α ⁺ MSCs. Graph in middle panel illustrates the flow cytometric quantification of CD31⁺ CD45⁺ Ter119⁺ Sca-1⁺ PDGFR α ⁺ MSCs in *Fbxw7^{ΔEC}* and littermate control femurs. Note increase of Sca-1⁺ and PDGFR α ⁺ MSCs in *Fbxw7^{ΔEC}* mutants ($n = 8$). Graph on right shows quantification of fibroblast colony-forming units (CFU-F) from *Fbxw7^{ΔEC}* and control femurs ($n = 7$). Data represent mean \pm s.e.m. P values, two-tailed unpaired t -test. Both assays confirm a significant increase in MSC frequency. **b**, Transcripts associated with osteogenic (*Sp7*), chondrogenic (*Acan*) and adipogenic (*Cfd*) differentiation were quantified by RT-qPCR after 14 days of differentiation culture of mesenchymal stromal cells isolated from femur. Data represent mean \pm s.d. ($n = 6$). P values, two-tailed unpaired t -test. No significant differences were seen for *Fbxw7^{ΔEC}* and control cells *in vitro*. The differentiation potential of *Fbxw7^{ΔEC}* cultured MSCs *ex vivo* was not altered. **c**, ELISA analysis of cellular SCF (lysate prepared from washed cells) in *Fbxw7^{ΔEC}* ($n = 8$ and 7) and *Rbpj^{ΔEC}* ($n = 7$ and 9) long bone. Data represent mean \pm s.d. P values, two-tailed unpaired t -test. **d**, Graphs showing ELISA analysis of extracellular SCF in *Fbxw7^{ΔEC}* ($n = 5$) and *Rbpj^{ΔEC}* ($n = 6$ and 5) long bone. Data represent mean \pm s.d. P values, two-tailed unpaired t -test. **e**, Increased chimaerism of *Fbxw7^{ΔEC}* mutant (also shown as Fig. 2f, upper panel) relative to littermate control BM is shown after primary and secondary transplantation (at 4 months after transplantation). Donor-derived chimaerism was analysed by transplanting BM cells harvested from *Fbxw7^{ΔEC}* mutant mice or littermate controls together with CD45.1 recipient-derived BM cells into lethally irradiated animals. For the secondary transplantation, 1×10^6 BM

cells from CD45.1 mice that had previously undergone transplantation at 1:1 ratio at 7 months after transplantation were injected into lethally irradiated recipients. Data represent mean \pm s.e.m. ($n = 5$ donors), P values, two-tailed unpaired t -test. **f**, Levels of donor-derived multi-lineage contribution were determined for *Fbxw7^{ΔEC}* and control BM cells at 18 weeks after transplantation by flow cytometry. HSC-CRU frequency and statistical significance were determined using ELDA software ($n = 3$ mice per dilution). Note significant increase in the HSC frequency in the *Fbxw7^{ΔEC}* mutant BM compared with littermate controls. **g**, Flow cytometric quantification of haematopoietic lineages in the *Fbxw7^{ΔEC}* and control BM. Data represent mean \pm s.e.m. ($n = 14$), P values, two-tailed unpaired t -test. Frequency of cells belonging to different haematopoietic lineages was not significantly altered in *Fbxw7^{ΔEC}* mutants. **h**, Representative confocal images of EC-specific *Notch2^{ΔEC}* or global *Notch4* mutants (*Notch4^{KO}*) and corresponding littermate control tibial bones after Emcn (red), PDGFR β (blue) and Sca-1 (green) immunostaining. **i**, Flow cytometric quantitation of type H ECs in *Notch2^{ΔEC}* and *Notch4* knockout bones relative to controls. Data represent mean \pm s.e.m. ($n = 5$ in two independent experiments), P values, two-tailed unpaired t -test. **j**, Flow cytometric quantitation of HSCs in *Notch2^{ΔEC}* and *Notch4* knockout BM relative to controls. Data represent mean \pm s.e.m. ($n = 5$ in two independent experiments), P values, two-tailed unpaired t -test. **k**, Representative confocal images of 4-week-old wild-type tibia showing Notch3 (green) and Emcn (red) immunostaining. Note absence of Notch3 expression in bone ECs. Nuclei, DAPI (blue). Unless otherwise mentioned, data presented in figure panels are based on three independent experiments.

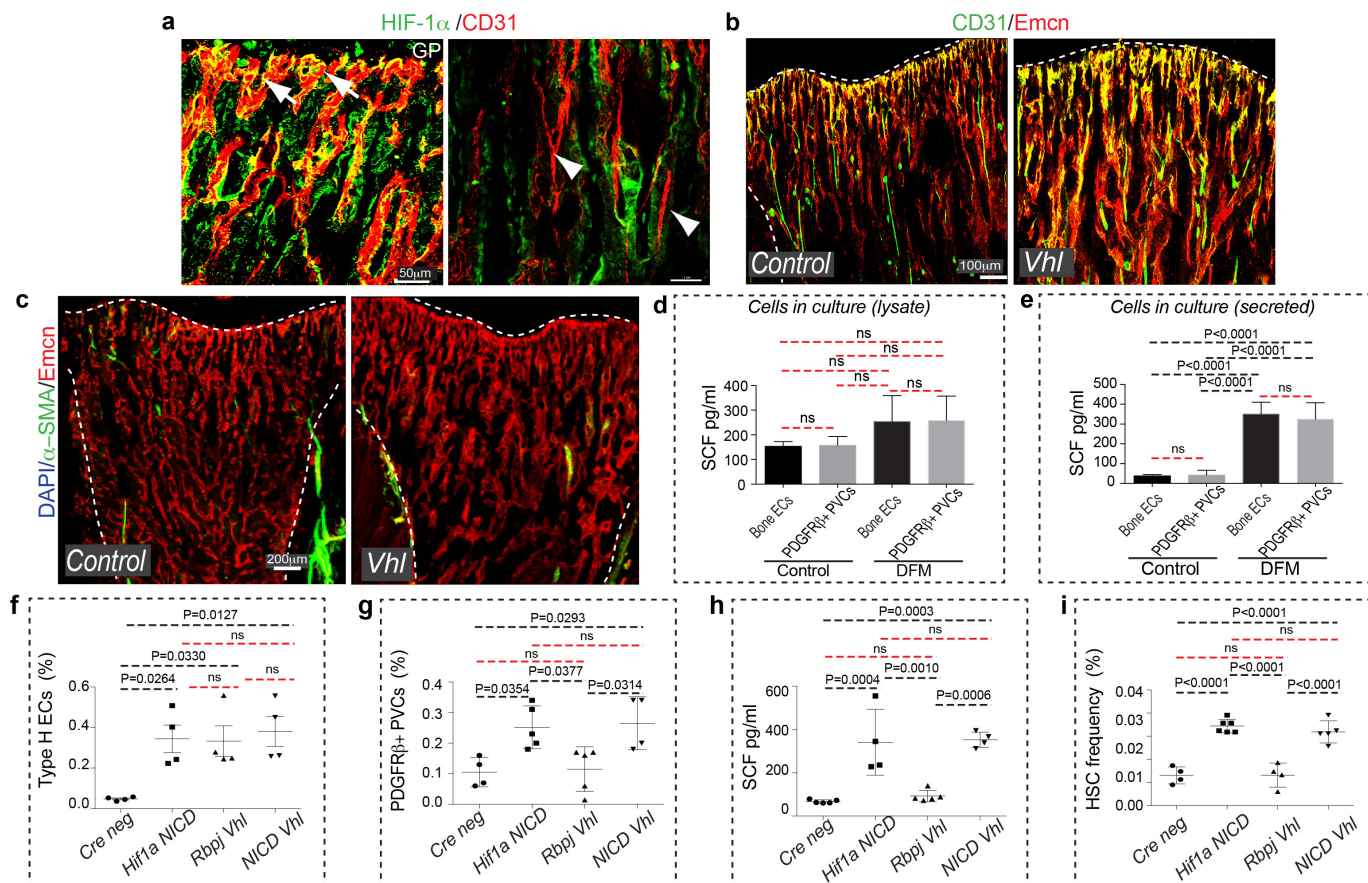


Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | HIF signalling in bone endothelium.

a, Maximum intensity projections of pimonidazole (green) stained young (8-week-old) and aged (45-week-old) tibia sections. Nuclei, DAPI (blue); CD31 immunostaining is shown in red. Dashed lines indicate the border of the growth plate. While pimonidazole staining was largely absent from 8-week-old metaphysis, hypoxic cells were readily detectable in the equivalent region in 45-week-old animals. **b**, Quantifications of *Hif1a* and *Epas1* transcripts in sorted ECs from young (4-week-old) and old (60-week-old) bone. Data represent mean \pm s.d. ($n = 11$). *P* values, two-tailed unpaired *t*-test. Endothelial expression of *Hif1a* transcripts was strongly reduced in aged animals, whereas expression of the related *Epas1* (*Hif2a*) gene was significantly increased. **c**, Quantitative analysis of *Cxcl12*, *Fgf1*, *Kitl*, *Tgfb3*, *Tgfb1* and *Vegfa* transcripts in dissected 60-week-old metaphysis relative to samples from young mice. Data represent mean \pm s.d. ($n = 5$). *P* values, two-tailed unpaired *t*-test. **d**, Phospho-MAPK (phosphoERK1/2; green) and Emcn (red) immunostaining in young (4-week-old) and aged (50-week-old) metaphysis. Nuclei, DAPI (blue). Dashed lines mark growth plate. **e**, Representative tile scan confocal images obtained from control, *Hif1a* ^{Δ EC} and double-mutant *Hif1a* ^{Δ EC} *Vhl* ^{Δ EC} tibial sections. Immunostaining for Emcn (red) and Osterix (green) is shown. Nuclei, DAPI (blue). The decline of type H ECs and Osterix⁺ cells in *Hif1a* ^{Δ EC} bone was not recovered in *Hif1a* ^{Δ EC} *Vhl* ^{Δ EC} double mutants. **f**, Graph showing *Fgf1*, *Pdgfa*, *Pdgfb*, *Tgfb1* and *Tgfb3* transcript levels in sorted ECs from *Hif1a* ^{Δ EC} and *Hif1a* ^{Δ EC} *Vhl* ^{Δ EC} double-mutant bones normalized to littermate control. Data

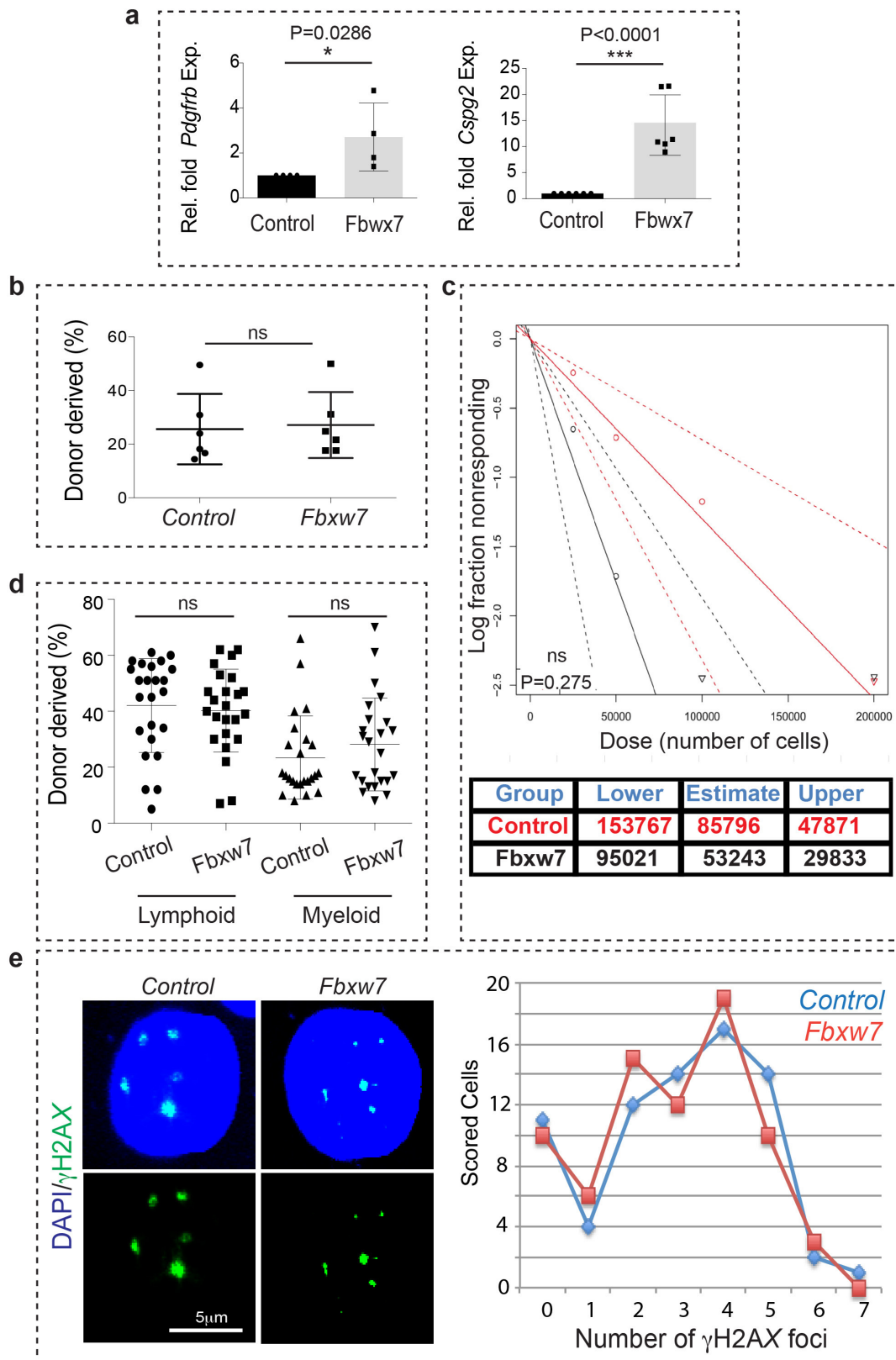
represent mean \pm s.d. ($n = 4-7$). *P* values, two-tailed unpaired *t*-test. **g**, Representative confocal images showing immunostaining for CD31 (green) and Emcn (red) in *Hif1a* ^{Δ EC} mutant, *Hif1a* ^{Δ EC} *Vhl* ^{Δ EC} double-mutant and control bones. **h**, Flow cytometric analysis of Sca1^{+/−} ($n = 5$) and ephrin-B2^{+/−} ECs ($n = 5$ and 6) among total CD45[−] Ter119[−] CD31⁺ ECs in *Hif1a* ^{Δ EC} long bones relative to Cre-negative littermates. Data represent mean \pm s.e.m. *P* values, two-tailed unpaired *t*-test. **i**, Representative confocal images of *Hif1a* ^{Δ EC} or control tibial metaphysis after Emcn (red) and PDGFR β (green) immunostaining. Nuclei, DAPI (blue). Dashed lines mark the growth plate. Note decline in Emcn^{hi} ECs and PDGFR β ⁺ perivascular cells in *Hif1a* ^{Δ EC} mutant. **j**, Analysis of donor-derived cells indicating LTR-HSC contribution, as determined 7 months after transplantation by flow cytometry. BM cells harvested from *Hif1a* ^{Δ EC} mutant mice or littermate controls were transplanted together with recipient CD45.1 recipient-derived BM cells into lethally irradiated recipient mice. Data represent mean \pm s.e.m. ($n = 6$ donors). *P* values, two-tailed unpaired *t*-test. **k**, Quantitative analysis of Sca1^{+/−} ECs ($n = 6$ and 5) and ephrin-B2^{+/−} ECs ($n = 5$ and 4) among total CD45[−] Ter119[−] CD31⁺ ECs in *Vhl* ^{Δ EC} long bone relative to Cre-negative littermates. Data represent mean \pm s.e.m. ($n = 4-6$). *P* values, two-tailed unpaired *t*-test. **l**, Representative confocal images of *Vhl* ^{Δ EC} and control tibial metaphysis with Emcn (red) and PDGFR β (green) immunostaining. Nuclei, DAPI (blue). Dashed lines mark growth plate. Unless otherwise mentioned, data presented in figure panels are based on three independent experiments.



Extended Data Figure 9 | Relation between Notch and HIF in bone ECs.

a, Metaphysis region of 2-week-old tibia after CD31 (red) and HIF-1 α (green) immunostaining. Dashed line marks the growth plate; arrows indicate type H endothelium. Note the presence of HIF-1 α in type H ECs (arrows) but the absence of signal in CD31⁺ arteries (arrowheads in right panel). **b**, Confocal images showing CD31 (green) and Emcn (red) immunostaining of *Vhl*^{iΔEC} mutant and littermate control tibia sections. Dashed lines mark growth plate. Note increase in type H capillaries. **c**, Tile scan confocal images showing α -SMA (green) and Emcn (red) immunostained tibia. Similar amounts of α -SMA⁺ cell-covered arteries were visible in *Vhl*^{iΔEC} and control samples. **d**, **e**, ELISA analysis of cellular (cell lysates, **d**) and secreted SCF levels (cell culture supernatant, **e**) in cell lysates of cultured BM-derived ECs and PDGFR β ⁺ perivascular cells (PVCs; $n = 5$ replicates) after treatment with vehicle control or DFM. Data represent mean \pm s.d. P values, one-way ANOVA with Tukey's multiple comparison post-hoc test. **f**, **g**, Frequency (percentage) of type H ECs (**f**, $n = 4$) and PDGFR β ⁺ PVCs (**g**, $n = 5$ mutants and 6 controls)

among total BM cells in *Hif1a*^{iΔEC} *NICD*^{iΔEC}, *Rbpj*^{iΔEC} *Vhl*^{iΔEC} and *NICD*^{iΔEC} *Vhl*^{iΔEC} double mutants relative to Cre-negative controls. Data represent mean \pm s.e.m. P values, one-way ANOVA with Tukey's multiple comparison post-hoc test. The combination of enhanced Notch and HIF activity in *NICD*^{iΔEC} *Vhl*^{iΔEC} double mutants failed to induce a bigger expansion of type H ECs. **h**, ELISA analysis of the cellular SCF levels in lysates of femur cells from *Hif1a*^{iΔEC} *NICD*^{iΔEC}, *Rbpj*^{iΔEC} *Vhl*^{iΔEC} and *NICD*^{iΔEC} *Vhl*^{iΔEC} double mutants relative to Cre-negative controls. Data represent mean \pm s.e.m. ($n = 4$ or 5 mutants and 5 controls). P values, one-way ANOVA with Tukey's multiple comparison post-hoc test. **i**, HSC frequency (%) in *Hif1a*^{iΔEC} *NICD*^{iΔEC}, *Rbpj*^{iΔEC} *Vhl*^{iΔEC} and *NICD*^{iΔEC} *Vhl*^{iΔEC} double mutants relative to Cre-negative controls. Data represent mean \pm s.e.m. ($n = 4$ or 5 mutants and 4 controls). P values, one-way ANOVA with Tukey's multiple comparison post-hoc test. Unless otherwise mentioned, data presented in figure panels are based on three independent experiments.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Properties of vascular niches and HSCs in aged mice. **a**, qPCR analysis of *Pdgfrb* and *Cspg4* expression (normalized to *Actb*) in long bone of *Fbxw7^{iΔEC}* mutants relative to littermate controls. Data represent mean \pm s.d. ($n = 4$ left panel; $n = 6$ right panel). P values, two-tailed unpaired t -test. **b**, Analysis of LTR-HSC contribution of BM cells from aged *Fbxw7^{iΔEC}* and control donors, as determined by flow cytometry at 16 weeks after competitive transplantation together with young CD45.1 BM cells (from 12- to 14-week-old mice) into lethally irradiated recipients. Data represent mean \pm s.e.m. ($n = 6$ donors). **c**, Levels of donor-derived multi-lineage contribution of aged *Fbxw7^{iΔEC}* and age-matched control BM cells as determined 18 weeks after transplantation by flow cytometry analysis. HSC-CRU frequency and

statistical significance were determined using ELDA software ($n = 3$ mice per dilution). **d**, Donor-derived lymphoid and myeloid contributions of aged *Fbxw7^{iΔEC}* and control BM cells determined 18 weeks after transplantation by flow cytometry analysis with B220 and CD11b antibodies. Data represent mean \pm s.d. ($n = 24$). P values, two-tailed unpaired t -test. **e**, Representative images and quantification of γ -H2AX foci in Lineage⁻/cKit⁺/Sca-1⁺ haematopoietic stem and progenitor cells (HSPCs) sorted from *Fbxw7^{iΔEC}* mice and littermate controls (110 HSPCs were scored for each group). Note persistence of γ -H2AX foci in the aged *Fbxw7^{iΔEC}* HSPCs. Unless otherwise mentioned, data presented in figure panels are based on three independent experiments.

The diversity-generating benefits of a prokaryotic adaptive immune system

Stineke van Houte¹, Alice K. E. Ekroth¹, Jenny M. Broniewski¹, Hélène Chabas^{1,2}, Ben Ashby^{3,4}, Joseph Bondy-Denomy⁵, Sylvain Gandon², Mike Boots^{3,4}, Steve Paterson⁶, Angus Buckling¹ & Edze R. Westra¹

Prokaryotic CRISPR-Cas adaptive immune systems insert spacers derived from viruses and other parasitic DNA elements into CRISPR loci to provide sequence-specific immunity^{1,2}. This frequently results in high within-population spacer diversity^{3–6}, but it is unclear if and why this is important. Here we show that, as a result of this spacer diversity, viruses can no longer evolve to overcome CRISPR-Cas by point mutation, which results in rapid virus extinction. This effect arises from synergy between spacer diversity and the high specificity of infection, which greatly increases overall population resistance. We propose that the resulting short-lived nature of CRISPR-dependent bacteria–virus coevolution has provided strong selection for the evolution of sophisticated virus-encoded anti-CRISPR mechanisms⁷.

We previously reported that *Pseudomonas aeruginosa* strain UCBPP-PA14 evolves high levels of CRISPR-Cas (clustered regularly interspaced short palindromic repeats; CRISPR-associated) adaptive immunity against virus DMS3vir under laboratory conditions⁶. However, viruses can readily evolve to overcome sequence-specific CRISPR immunity^{8,9}. To study how CRISPR-Cas impacts virus persistence, we measured titres of virus DMS3vir over time upon infection of either wild-type *P. aeruginosa* or a functional CRISPR-Cas knockout (CRISPR KO) strain. Virus that infected the wild-type strain went extinct at 5 days post-infection (d.p.i.) (Fig. 1a), whereas virus infecting the CRISPR KO strain persisted in all replicates until the experiment was terminated at 30 d.p.i. (Fig. 1b). Wild-type bacteria exclusively evolved CRISPR-mediated immunity, while the CRISPR KO strain evolved immunity by mutation, loss or masking of the receptor (that is, surface mutation) (Extended Data Fig. 1a). The observation that CRISPR-Cas drives virus extinct so rapidly was unexpected since viruses can escape CRISPR immunity by a single point mutation^{8,9}.

Virus extinction might result from the high level of spacer diversity that naturally evolves upon virus exposure in this and other CRISPR-Cas systems³⁻⁶. Both theory and data suggest that host genetic diversity can synergistically reduce the spread of parasites if the infection process is specific (that is, a parasite genotype can infect a restricted number of host genotypes) and a failed infection results in parasite death¹⁰⁻¹⁸; assumptions that hold for CRISPR-Cas-virus interactions. While the protective effect of host diversity may be lost following the evolution of single viruses that escape from multiple spacers^{10,17}, host diversity has the additional benefit of limiting such viral adaptation. Specifically, lower virus population sizes resulting from host diversity^{11,12} reduces the probability of escape mutations, and the greater the diversity the more escape mutations needed.

To examine these hypotheses, we generated bacterial populations in which we manipulated the level of spacer diversity; we used 48 individual clones with CRISPR-based immunity against virus DMS3vir to generate bacterial populations with five distinct diversity levels: monocultures or polycultures consisting of equal mixtures of either

6, 12, 24 or 48 clones. To allow for direct comparisons, each of the 48 clones was equally represented at each diversity level by adjusting the number of replicate experiments accordingly. Each population was competed against a previously described surface mutant⁶ in the presence or absence of virus DMS3vir and virus levels were monitored over time.

This experiment revealed a strong inverse relationship between virus persistence and the level of spacer diversity in the bacterial population (Fig. 2). Virus titres remained high in 44 out of 48 replicates when the CRISPR population consisted of a monoculture (Fig. 2a). However, as diversity increased, virus persistence decreased (Fig. 2b–e) and virus was driven extinct rapidly and reproducibly when the CRISPR population consisted of a 48-clone mixture (Fig. 2e).

Next, we examined the fitness consequences of generating spacer diversity. In the absence of virus there was no significant effect of diversity on the relative fitness associated with CRISPR-Cas compared with a resistant surface mutant (Extended Data Fig. 1b; $F_{1,52} = 3.20$, $P = 0.08$). However, in the presence of virus, CRISPR-associated fitness increased with increasing spacer diversity (Fig. 3; $F_{4,71} = 40.30$, $P < 0.0001$ and Extended Data Table 1), with mean fitness increasing 11-fold from monoculture to the highest diversity population. In monoculture, the CRISPR population was outcompeted by the surface mutant (relative fitness < 1 ; $T = -11.68$, $P < 0.0001$). However, as diversity increased, the CRISPR population consistently outcompeted the surface mutant (relative fitness > 1 ; 6 clones: $T = 3.05$, $P = 0.0093$; 12 clones: $T = 3.95$, $P = 0.0028$; 24 clones: $T = 3.48$, $P = 0.0088$; 48 clones: $T = 3.06$, $P = 0.014$; all significant after sequential Bonferroni correction¹⁹), showing that the generation of spacer diversity is an important fitness determinant of CRISPR-Cas (Fig. 3).

Given that all bacterial clones used in the experiment were initially resistant, we hypothesized that the benefit of spacer diversity emerges from an inability of virus to evolve escape mutants. To examine this,

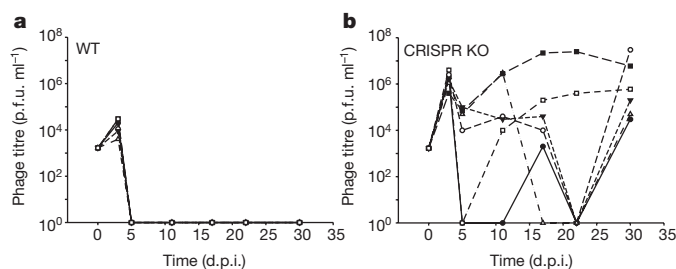


Figure 1 | Evolution of CRISPR-mediated immunity leads to rapid extinction of virus. **a, b,** Titre (in plaque-forming units per millilitre) of virus DMS3vir over time upon infection of wild-type (WT) *P. aeruginosa* (**a**) and *P. aeruginosa* strain *csy3::LacZ* (**b**) (CRISPR KO strain). Each line indicates a biological replicate experiment ($n = 6$). The limit of detection is 200 p.f.u. ml⁻¹.

¹ESI and CEC, Biosciences, University of Exeter, Cornwall Campus, Penryn TR10 9EZ, UK. ²CEFE UMR 5175, CNRS-Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, 1919, route de Mende 34293, Montpellier Cedex 5, France. ³Department of Integrative Biology, University of California, Berkeley, California 94720, USA. ⁴CEC, Biosciences, University of Exeter, Cornwall Campus, Penryn TR10 9EZ, UK. ⁵Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, California 94158, USA. ⁶Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK.

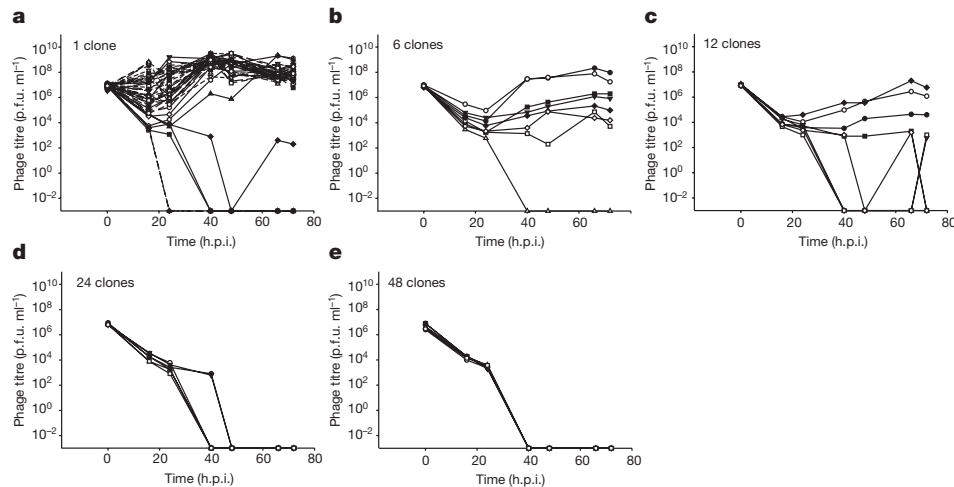


Figure 2 | Virus persistence inversely correlates with the level of spacer diversity. a–e, Virus titres (in plaque-forming units per millilitre) over time upon infection of a bacterial population consisting of an equal mixture of a surface mutant and a monoculture with CRISPR-mediated immunity (a; $n = 48$), or polycultures with CRISPR-mediated immunity

consisting of 6 clones (b; $n = 8$), 12 clones (c; $n = 8$), 24 clones (d; $n = 6$), 48 clones (e; $n = 6$). The number of replicates is chosen such that all clones are equally represented in each treatment. Each line indicates a biological replicate experiment; in all cases n is the number of biological replicates. The limit of detection is 200 p.f.u. ml⁻¹.

virus isolated from each time point (0, 16, 24, 40, 48, 64 and 72 hours post-infection (h.p.i.)) was spotted onto lawns of each of the 48 CRISPR clones. As expected, we could not detect escape virus in the ancestral virus (Fig. 4a; left column, indicated in green). However, in 43 of the 48 CRISPR monocultures, virus evolved within 2 days to overcome CRISPR immunity (Fig. 4a, indicated in red). For five clones no escape virus could be detected, and virus became extinct in four of these instances (Fig. 4a, asterisks). Three of these five clones carried multiple spacers targeting the virus, which limits the emergence of escape virus¹⁶. The emergence of escape virus decreased as diversity increased to 6, 12, 24 and 48 CRISPR alleles (Fig. 4); in the last two, no escape virus could be detected. These phenotypic data were supported by results of deep sequencing of virus genotypes isolated from 1 d.p.i.: there was a significant inverse relationship between host diversity and the accumulation of viral mutations in the target sequences (Extended Data Fig. 1c, d). This is because virus needs to overcome multiple spacers in the diverse host population if it is to increase in frequency (Extended Data Fig. 1e, f). Consistent with a lack of escape virus emerging against all host genotypes, the spacer content of mixed populations of 6, 12, 24 and 48 clones did not increase between $t = 0$ and $t = 3$ d.p.i. (Wilcoxon signed rank $P > 0.2$ for all treatments), whereas monocultures acquired novel spacers in response to emerging escape virus (Wilcoxon signed rank $W = 333$, d.f. = 47, $P < 0.0001$; Extended Data Fig. 1g). These data show that although escape viruses can clearly evolve against most of the clones, they do not emerge when these clones are mixed.

We hypothesized that the benefit of within-population spacer diversity is due to synergy between the different clones. However, diversity will also increase the chance that the population will contain a single clone with one or more spacers that the virus is unable to overcome. Indeed, we observed five clones against which escape virus mutants were never detected, and presence of these clones in many of the diverse populations could explain the fitness advantage of diversity. To investigate if synergy plays an important role in the benefit of diversity beyond this 'jackpot' effect, we compared the fitness of diverse populations with the fitness of the fittest constituent clone, as measured in monoculture. This analysis revealed that synergism contributed an approximately 50% growth rate advantage when in competition with surface mutants (mean \pm s.e.m. difference in fitness between mixtures and fittest constituent monoculture = 0.47 ± 0.18 ; $P < 0.01$).

The short-lived nature of coevolution between CRISPR-resistant bacteria and virus escape mutants beyond a host diversity threshold may explain the evolution of sophisticated anti-CRISPR mechanisms

to overcome CRISPR-Cas⁷. Indeed, the same DMS3vir virus carrying an anti-CRISPR gene⁷ was found to persist independent of CRISPR diversity levels (Extended Data Fig. 1h, i) and caused similar extinction of CRISPR-resistant monocultures and 48-clone populations that competed against a surface mutant (Fisher's exact test, $P = 1.0$ at $t = 1$ d.p.i., $P = 0.33$ at $t = 3$ d.p.i.; Extended Data Fig. 1j).

Finally, to test that our results were not limited to the *P. aeruginosa* PA14 type I-F CRISPR-Cas system, we performed a similar experiment with *Streptococcus thermophilus* DGCC7710 clones that evolved resistance against virus 2972 using a type II-A CRISPR-Cas system. As shown in Extended Data Fig. 2, we found a similar effect of CRISPR resistance allele diversity on virus persistence and escape virus emergence. However, during coevolution experiments the levels of evolved spacer diversity are lower in *S. thermophilus* (data not shown and refs 4, 5), which, consistent with theory^{10,17}, allows for more persistent coevolution^{4,5}. Lower levels of evolved spacer diversity might be due to a more weakly primed CRISPR-Cas system^{20–22}.

Collectively, our data demonstrate that the propensity to generate host genetic diversity is a key fitness determinant of CRISPR-Cas adaptive immune systems because it limits the emergence of escape virus. Consistent with the idea that it is harder for a parasite to adapt to a heterogeneous host population²³, virus rapidly evolved high levels of

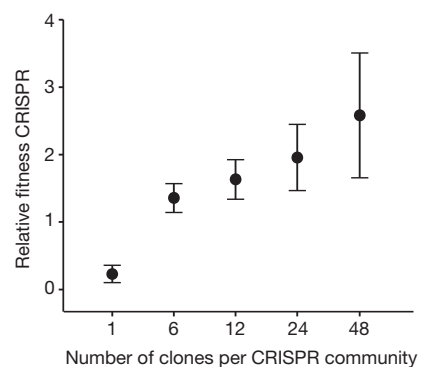


Figure 3 | The benefit of CRISPR immunity increases with increasing spacer diversity. Relative fitness of bacterial populations with CRISPR-mediated immunity, with spacer diversity as indicated, at 3 d.p.i. when competing with a surface mutant. Error bars, 95% CI. In all cases, the number of biological replicates equals the n values given in the legend of Fig. 2.

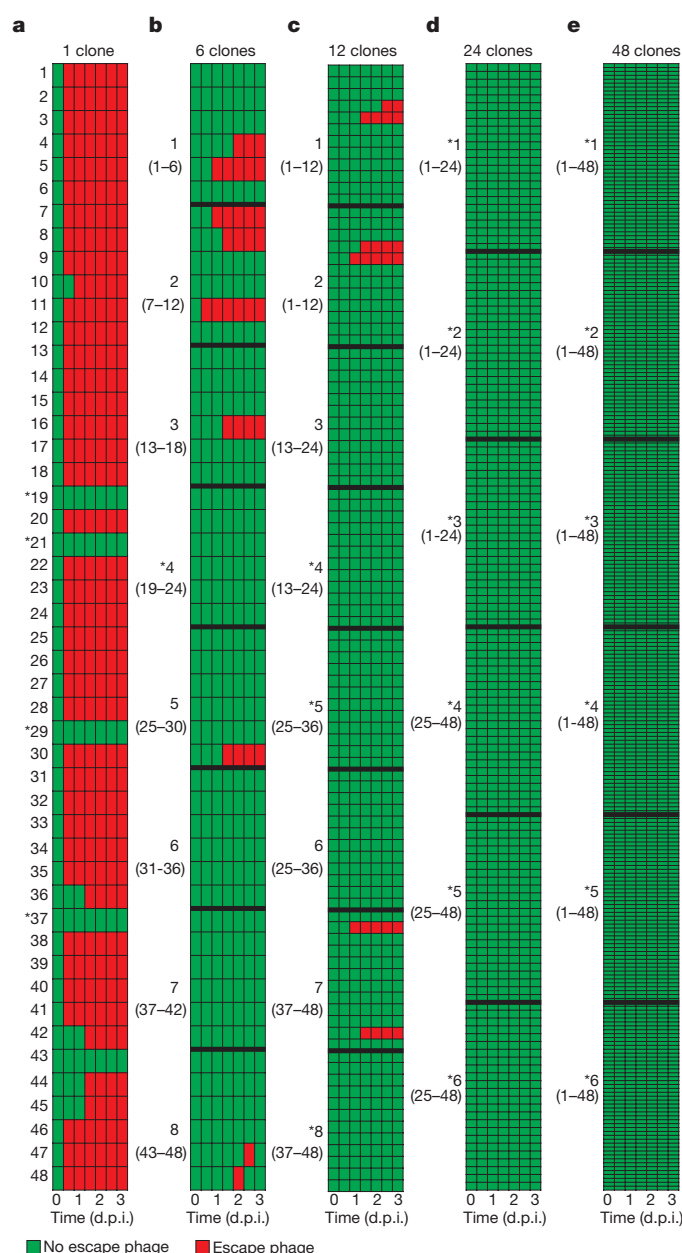


Figure 4 | Evolution of virus infectivity is constrained by spacer diversity. Emergence of virus that overcomes host CRISPR immunity (escape virus) during the experiment shown in Fig. 2 (a–e correspond to Fig. 2a–e). Table columns correspond to time points where virus was isolated (0, 16, 24, 40, 48, 64 and 72 h.p.i.; indicated below the table in days post-infection). Green, no escape virus; red, escape virus. Bold numbers indicate individual biological replicates, as detailed in the legend of Fig. 2. In b–e replicates are separated by bold lines in the table. Numbers between parentheses refer to the clones in the CRISPR population. Asterisks indicate virus extinction.

infectivity on monocultures, but not on a diverse mix of the same host genotypes. Parasites are often invoked as the selective force driving the evolution of diversity generating mechanisms^{23–26}. In most cases, individual-level selection is assumed to be the driver of these traits, because individual benefits are high, and group selective benefits would be opposed by the invasion of individuals who do not pay the fitness costs associated with these mechanisms (for example, sex and increased mutation rates)^{26–28}. In the case of CRISPR-Cas, we speculate that population-level selection may have contributed to its evolution. First, there were large benefits associated with synergy between diverse genotypes. Second, costs of CRISPR-Cas are conditional on virus exposure^{6,29},

and clones lacking CRISPR immunity cannot invade populations (Extended Data Figs 3 and 4). Third, the highly structured nature of bacterial populations, and the resulting high relatedness, promotes between-population selection³⁰. Future tests of this hypothesis are needed to reconcile the selective forces that have shaped the evolution of CRISPR-Cas systems.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 September 2015; accepted 19 February 2016.

Published online 13 April 2016.

- van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Rev. Microbiol.* **12**, 479–492 (2014).
- Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Andersson, A. F. & Banfield, J. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050 (2008).
- Paez-Espino, D. *et al.* Strong bias in bacterial CRISPR elements that confer immunity to phage. *Nature Commun.* **4**, 1430 (2013).
- Paez-Espino, D. *et al.* CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio* **6**, e00262–e15 (2015).
- Westra, E. R. *et al.* Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* **25**, 1043–1049 (2015).
- Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
- Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
- Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeats (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108**, 10098–10103 (2011).
- Childs, L. M., England, W. E., Young, M. J., Weitz, J. S. & Whitaker, R. J. CRISPR-induced distributed immunity in microbial populations. *PLoS ONE* **9**, e101710 (2014).
- Lively, C. M. The effect of host genetic diversity on disease spread. *Am. Nat.* **175**, E149–E152 (2010).
- King, K. C. & Lively, C. M. Does genetic diversity limit disease spread in natural populations? *Heredity* **109**, 199–203 (2012).
- Van Baalen, M. & Beekman, M. The costs and benefits of genetic heterogeneity in resistance against parasites in social insects. *Am. Nat.* **167**, 568–577 (2006).
- Altarmatt, F. & Ebert, D. Genetic diversity of *Daphnia magna* populations enhances resistance to parasites. *Ecol. Lett.* **11**, 918–928 (2008).
- Schmid-Hempel, P. & Crozier, R. H. Polygyny versus polyandry versus parasites. *Phil. Trans. R. Soc. Lond. B* **354**, 507–515 (1999).
- Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The population and evolutionary dynamics of phage and bacteria with CRISPR immunity. *PLoS Genet.* **9**, e1003312 (2013).
- Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Evolutionary dynamics of the prokaryotic adaptive immune system CRISPR-Cas in an explicit ecological context. *J. Bacteriol.* **195**, 3834–3844 (2013).
- Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).
- Rice, W. R. Analyzing tables of statistical tests. *Evolution* **43**, 223–225 (1989).
- Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Commun.* **3**, 945 (2012).
- Swarts, D. C., Mosterd, C. van Passel, M. W. & Brouns, S. J. CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* **7**, e35888 (2012).
- Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl Acad. Sci. USA* **111**, E1629–E1638 (2014).
- Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites (a review). *Proc. Natl Acad. Sci. USA* **87**, 3566–3573 (1990).
- Pal, C., Maciá, M. D., Oliver, A., Schachar, I. & Buckling, A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**, 1079–1081 (2007).
- Morran, L. T., Schmidt, O. G., Gelarden, I. A., Parrish, R. C., II & Lively, C. M. Running with the red queen: host-parasite coevolution selects for biparental sex. *Science* **333**, 216–218 (2011).
- Ashby, B. & King, K. C. Diversity and the maintenance of sex by parasites. *J. Evol. Biol.* **28**, 511–520 (2015).
- Lively, C. M. An epidemiological model of host–parasite coevolution and sex. *J. Evol. Biol.* **23**, 1490–1497 (2010).
- Peters, A. D. & Lively, C. The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. *Am. Nat.* **154**, 393–405 (1999).
- Vale, P. F. *et al.* Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proc. R. Soc. B* **282**, 20151270 (2015).
- Gardner, A. & Grafen, A. Capturing the superorganism: a formal theory of group adaptation. *J. Evol. Biol.* **22**, 659–671 (2009).

Acknowledgements We thank D. Morley and S. Kay for experimental contributions and A. Gardner for comments on the manuscript. S.v.H. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 660039. E.R.W. received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under Research Executive Agency grant agreement number 327606. E.R.W., A.B. and M.B. also acknowledge the Natural Environment Research Council, the Biotechnology and Biological Sciences Research Council, the Royal Society, the Leverhulme Trust, the Wellcome Trust and the AXA research fund for funding. J.M.B.-D. was supported by the University of California San Francisco Program for Breakthrough in Biomedical Research, the Sandler Foundation, and a National Institutes of Health Director's Early Independence Award (DP5-OD021344). H.C. was funded by the Erasmus+ programme (European Union), the Explora'Sup programme (Région Rhône-Alpes) and the Centre Régional des Œuvres Universitaires et Scolaires (CROUS; French State).

Author Contributions E.R.W., A.B. and S.v.H. conceived and designed the experiments. H.C. performed coevolution experiments. S.v.H., E.R.W., A.K.E.E. and J.M.B. performed all competition experiments and associated analysis of virus persistence and host and virus evolution. S.P. performed and analysed deep sequencing of virus genomes. J.B.-D. supplied virus with anti-CRISPR gene. B.A. and M.B. contributed to discussions and provided feedback throughout the project. S.G. and H.C. helped to set up the experiments with *S. thermophilus*. S.v.H., E.R.W. and A.B. wrote the manuscript.

Author Information Sequence data are available from the European Nucleotide Archive under accession number PRJEB12001 and analysis scripts are available from <https://github.com/scottishwormboy/vanHoute>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.R.W. (westra.edze@gmail.com), S.v.H. (vanhoute.stineke@gmail.com) or A.B. (A.J.Buckling@exeter.ac.uk).

METHODS

The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Bacterial strains and viruses. *P. aeruginosa* UCBPP-PA14 (WT), *P. aeruginosa* UCBPP-PA14 *csy3::LacZ* (referred to as CRISPR KO, which carries a disruption of an essential *cas* gene and can therefore not evolve CRISPR immunity), the CRISPR-KO-derived surface mutant and virus DMS3vir have all been described in ref. 6 and references therein. Phage DMS3vir+*acrF1*, which carries the anti-CRISPR gene *acrF1* (formerly 30–35), was made by inserting *acrF1* into the DMS3vir genome using methods described in ref. 7. *S. thermophilus* strain DGCC7710 and its virus 2972 have been described in ref. 2.

Coevolution experiments. The coevolution experiments shown in Fig. 1 were performed in glass microcosms by inoculating 6 ml M9 supplemented with 0.2% glucose with approximately 10^6 colony-forming units of bacteria from fresh overnight cultures of the wild-type *P. aeruginosa* UCBPP-PA14 or CRISPR KO strain and adding 10^4 p.f.u. of virus DMS3vir, followed by incubation at 37 °C while shaking at 180 r.p.m. (six replicates). Cultures were transferred daily 1:100 to fresh broth. Virus titres were determined at 0, 3, 5, 11, 17, 22 and 30 days after the start of the coevolution experiment by spotting virus samples isolated by chloroform extraction on a lawn of CRISPR KO bacteria. The analysis of virus immunity was performed by cross-streak assay and PCR as described previously⁶.

Generation of populations with different levels of CRISPR diversity. For the competition experiments, shown in Figs 2–4 and Extended Data Figs 1b–j, 3 and 4, we generated *P. aeruginosa* populations with varying levels of CRISPR spacer (allele) diversity. To this end, we isolated from the six replicates of the coevolution experiment (Fig. 1) a total of 48 individual clones that had acquired CRISPR immunity against virus DMS3vir. We have previously shown that individual clones tend to have unique spacers⁶. Using these 48 clones, populations with five different levels of CRISPR spacer (allele) diversity were generated. These populations consisted of (1) 1 clone (a monoculture; a clonal population carrying a single spacer); equal mixtures of (2) 6 clones; (3) 12 clones; (4) 24 clones and (5) 48 clones. In total, 48 different monocultures (48 × monocultures), 8 × 6-clone populations, 4 × 12-clone populations, 2 × 24-clone populations and 1 × 48-clone population were generated (details of the composition of each population can be found below, in the section ‘Number of replicate experiments’).

Competition experiments. Competition experiments were done in glass microcosms in a total volume of 6 ml M9 supplemented with 0.2% glucose. Competition experiments were initiated by inoculating 1:100 from a 1:1 mixture (in M9 salts) of overnight cultures of the appropriate CRISPR population and either the surface mutant (Figs 2–4 and Extended Data Fig. 1b–j) or the CRISPR KO strain (Extended Data Figs 3 and 4). At the start of each experiment 10^9 p.f.u. of virus was added, unless indicated otherwise. Cultures were transferred daily 1:100 into fresh broth. At 0 and 72 h.p.i. samples were taken and cells were serially diluted in M9 salts and plated on LB agar supplemented with 50 µg ml^{−1} X-gal (to allow discrimination between wild-type-derived CRISPR clones (white) and CRISPR KO or surface mutant (blue)). The relative frequencies of the wild-type strain were used to calculate the relative fitness (relative fitness = [(fraction strain A at $t = x$) × (1 – (fraction strain A at $t = 0$))]/[(fraction strain A at $t = 0$) × (1 – (fraction strain A at $t = x$))]). At 0, 16, 24, 40, 48, 66 and 72 h.p.i., samples were taken and chloroform extractions were performed to isolate total virus, which was spotted on a lawn of CRISPR KO bacteria for quantification. All subsequent statistical analyses were performed using JMP (version 12) software.

Determination of escape virus emergence. To determine the emergence of escape virus during the competition experiments, every isolated virus sample was spotted onto 48 different bacterial lawns, corresponding to each of the different CRISPR clones. This procedure was done for each of the seven time points (see above), to enable us to track the emergence of escape virus against every individual clone over the time course of the experiment.

Deep sequencing. Isolated phage samples from $t = 1$ d.p.i. of the competition experiment shown in Figs 2–4 were used to perform deep sequencing of spacer target sites on the phage genomes. To obtain sufficient material, phage were amplified by plaque assay on the CRISPR KO strain. Viruses from all replicates within a single diversity treatment were pooled. As a control, ancestral virus and escape virus from competition between surface mutant and monocultures of CRISPR clones 1–3 were processed in parallel. Virus genomic DNA extraction was performed from 5 ml sample at approximately 10^{10} p.f.u. ml^{−1} using the Norgen phage DNA isolation kit, following the manufacturer’s instructions. Barcoded Illumina Truseq Nano libraries were constructed from each DNA sample with an

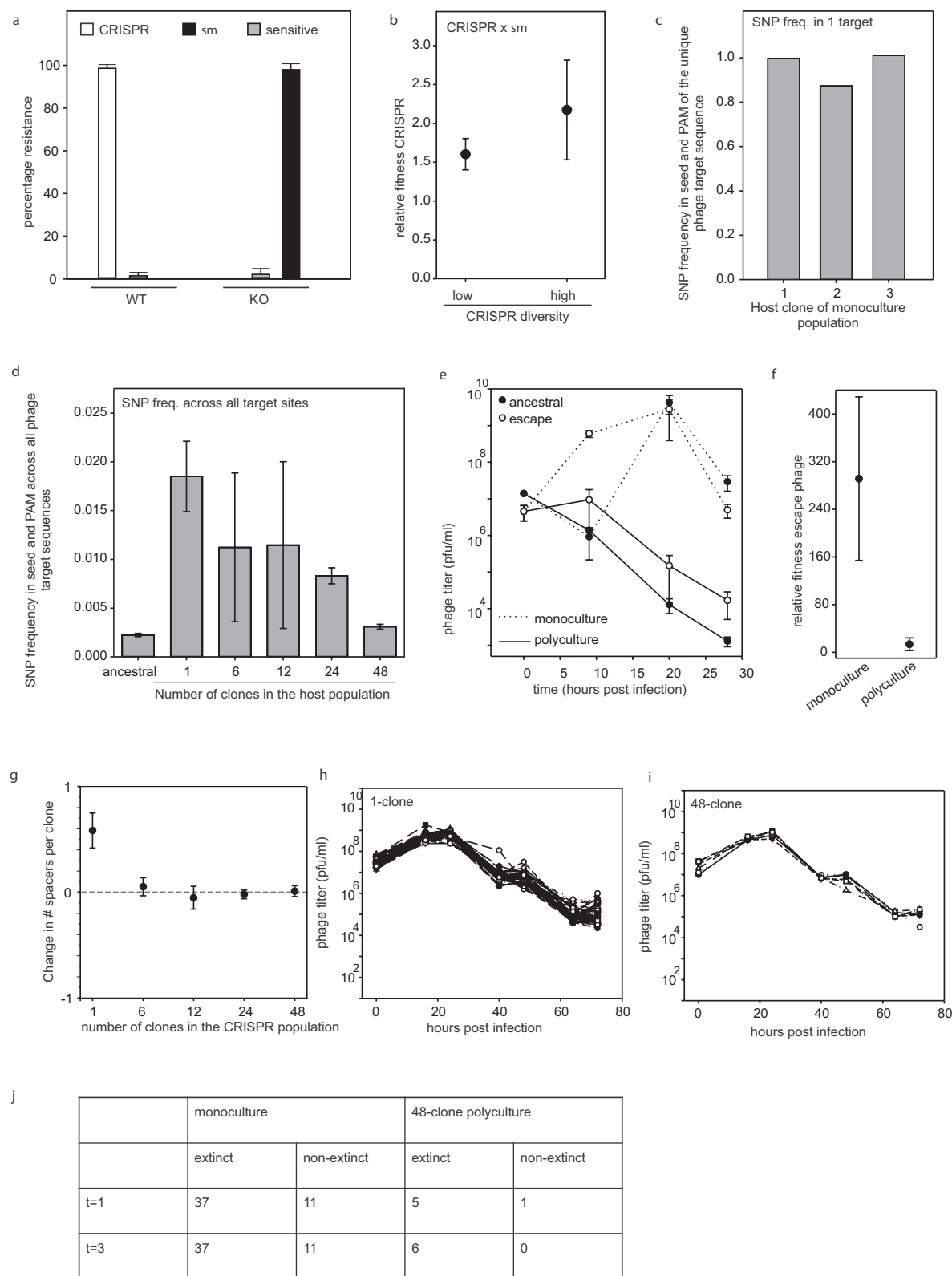
approximately 350 bp insert size and 2 × 250 bp reads generated on an Illumina MiSeq platform. Reads were trimmed using Cutadapt version 1.2.1 and Sickle version 1.200 and then overlapping reads merged using Flash version 1.2.8 to create high quality sequence at approximately 8,000 × coverage of DMS3vir per sample. These reads were mapped to PA14 and DMS3vir genomes using bwa mem version 0.7.12 and allele frequencies of single nucleotide polymorphisms within viral target regions quantified using samtools mpileup version 0.1.18. Further statistical analyses was performed in R version 3.2.2. Sequence data are available from the European Nucleotide Archive under accession number PRJEB12001 and analysis scripts are available from <https://github.com/scottishwormboy/vanHoute>.

Determining the acquisition of new spacers. To examine spacer acquisition during the competition experiments shown in Figs 2–4, we examined by PCR for each diversity treatment the spacer content of 384 randomly isolated clones at both $t = 0$ and $t = 3$ (192 clones per time point). For each replicate experiment, the difference in the total number of spacers between $t = 0$ and $t = 3$ was divided by the number of clones that were examined to calculate the average change in the number of spacers per clone.

Number of replicate experiments. To ensure equal representation of each of the 48 clones across the different treatments, the number of replicate experiments for a given diversity treatment was adjusted accordingly, with a total number of replicates of at least six for sufficient statistical power. Hence, competition experiments with the one-clone (monoculture) populations were performed in 48 independent replicates, each corresponding to a unique monoculture of a CRISPR clone (clones 1–48; each clone is equally represented). Competition experiments with the 6-clone populations were performed in eight independent replicates, each corresponding to a unique polyculture population (population 1: equal mixture of clones 1–6; population 2: clones 7–12; population 3: clones 13–18; population 4: clones 19–24; population 5: clones 25–30; population 6: clones 31–36; population 7: clones 37–42; population 8: clones 43–48). Competition experiments with the 12-clone populations were also performed in eight replicates, corresponding to four unique polyculture populations (replicate 1 and 2: clones 1–12; replicate 3 and 4: clones 13–24; replicate 5 and 6: clones 25–36; replicate 7 and 8: clones 37–48). Competition experiments with the 24-clone populations were performed in six replicates, corresponding to two unique polyculture populations (replicate 1–3: clones 1–24; replicate 4–6: clones 25–48). Competition experiments with the 48-clone populations were performed in six replicates, each corresponding to the same polyculture population (replicate 1–6: clones 1–48).

Escape phage degradation and fitness. In the experiment shown in Extended Data Fig. 1e, f, approximately 10^8 p.f.u. of either ancestral virus or escape virus, which was isolated from the competitions between monocultures 1–6 and the surface mutant, were used to infect a monoculture of the corresponding CRISPR clone or the 48-clone polyculture. Phage samples were taken at 0, 9, 20 and 28 h.p.i. by chloroform extraction and titrated on a lawn of the CRISPR KO strain. Fitness of each of the escape phages was determined by a competition experiment between ancestral and escape virus; a 50:50 ratio of escape and ancestral phage (10^9 p.f.u. total) was used to infect either a monoculture of the corresponding CRISPR clone or the 48-clone polyculture. Virus samples were taken at $t = 0$ and $t = 20$ h.p.i. by chloroform extraction and used in a plaque assay on CRISPR KO. Next, individual plaques (48 plaques per replicate) were isolated and amplified on the CRISPR KO strain. To determine the ratio of escape and ancestral virus, virus from each individual plaque was spotted on a lawn of (1) CRISPR KO (both ancestral and escape virus form plaques) and (2) the corresponding CRISPR immune clone (only escape virus can form a plaque).

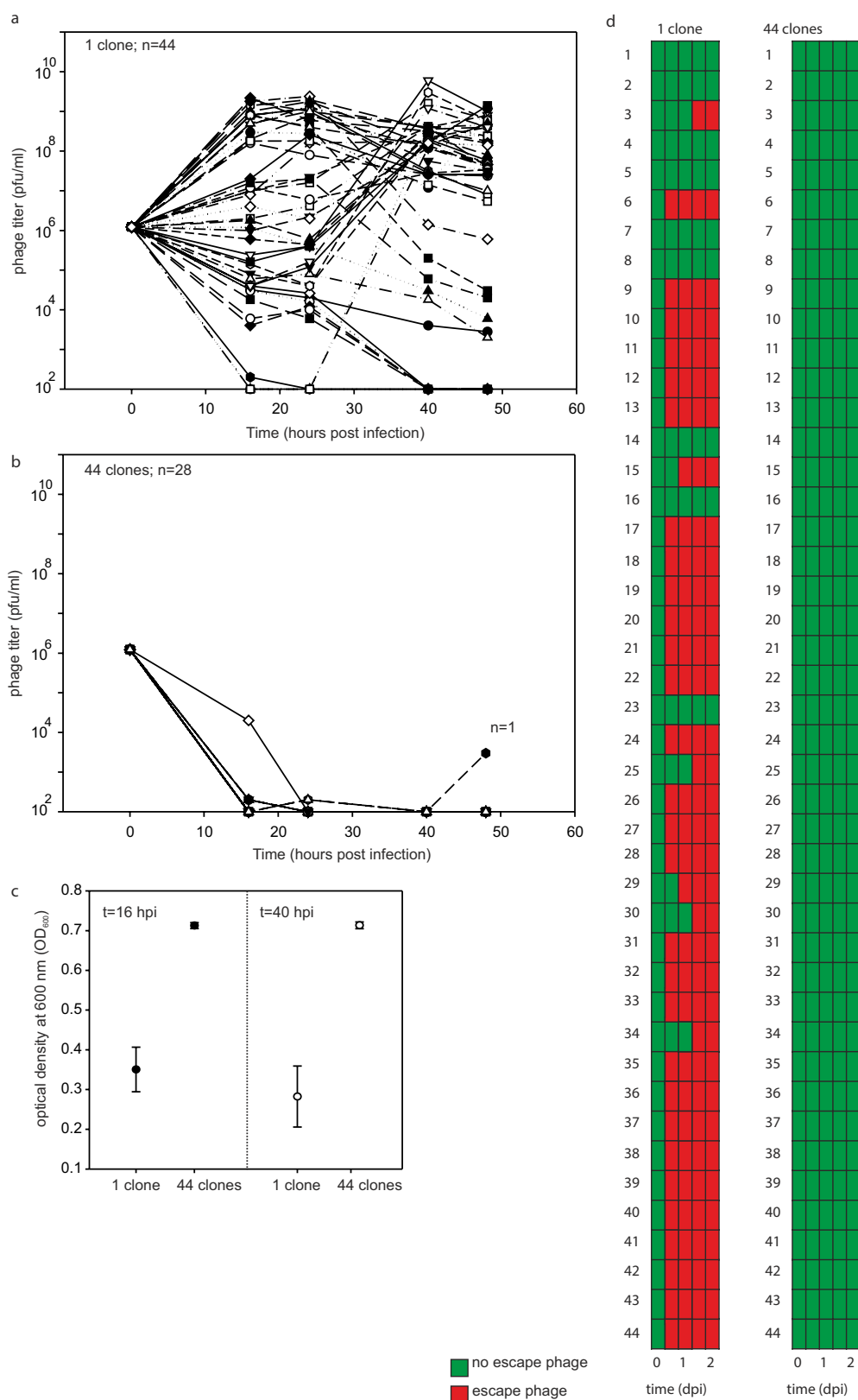
Effect of spacer diversity in *S. thermophilus*. *S. thermophilus* DGCC7710 was grown in M17 medium supplemented with 0.5% α-lactose (LM17) at 42 °C. Virus 2972 was used throughout the experiments. Virus infections were performed using 10^6 p.f.u. of phage 2972 and 10 mM CaCl₂ to facilitate the infection process. To obtain virus-resistant *S. thermophilus* clones, a sample of virus lysate at 24 h.p.i. was plated on LM17 agar plates. Individual colonies were picked and PCR-screened for the acquisition of novel spacers in each of the four CRISPR loci, as described in ref. 2. A total of 44 individual clones with a novel spacer in CRISPR1 (see ref. 16 and references therein) were selected to generate 44 monocultures and a single polyculture comprising a mix of 44 clones. These cultures were infected with 10^7 p.f.u. of virus, and samples were taken after the indicated periods of time to isolate virus. We determined virus titres by spotting viral dilutions on lawns of ancestral bacteria, and the emergence of escape virus by spotting virus on lawns corresponding to each of the 44 CRISPR-resistant clones.



Extended Data Figure 1 | See next page for caption.

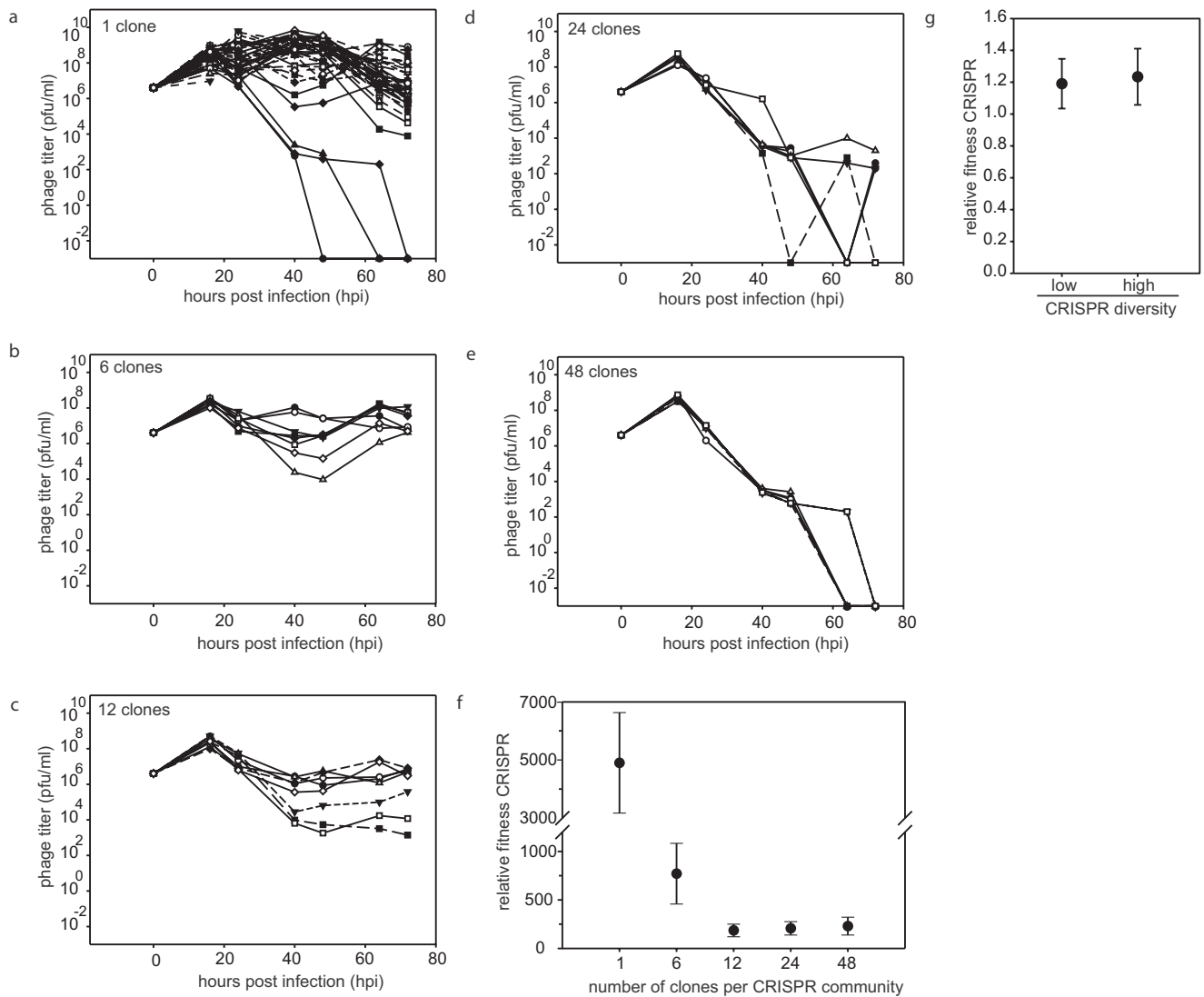
Extended Data Figure 1 | CRISPR diversity drives virus extinct since virus cannot escape by point mutation. **a**, Percentage bacteria (WT or CRISPR KO) from the experiment shown in Fig. 1 that have evolved CRISPR immunity (white bar), surface immunity (black bar) or remained sensitive (sensitive; grey bars) at 5 d.p.i. with virus DMS3vir ($n = 6$ for both treatments). **b**, Relative fitness of CRISPR immune monocultures (single spacer; low diversity, $n = 48$) and polycultures (48 spacers; high diversity, $n = 6$) at 3 d.p.i. when competing with a surface mutant (*sm*) in the absence of virus. **c**, **d**, Deep sequencing analysis of the frequency of mutations in the seed region and PAM of the target sequence of virus isolated at $t = 1$ d.p.i. from the experiment shown in Fig. 4. **c**, Frequency of point mutation in the target sequence of viral populations isolated from monoculture $1-3 \times$ surface mutant competitions. **d**, Average frequency of point mutation across all target sites in the ancestral virus genome and in the genomes of virus from pooled samples of all biological replicates from a single diversity treatment (monocultures: $n = 48$; 6-clone: $n = 8$; 12-clone: $n = 8$; 24-clone: $n = 6$; 48-clone: $n = 6$). **e**, **f**, Virus that escapes a single spacer present in a diverse CRISPR population decreases in frequency, despite a fitness benefit over ancestral virus. **e**, Titres (in plaque-forming units per millilitre) over time upon infection of monocultures (dotted line) or polycultures of 48 clones (solid line) with approximately 10^8 p.f.u. ancestral (closed circles) or escape (open circles) virus. Escape virus was isolated from monocultures

$1-6 \times$ surface mutant competitions shown in Figs 2–4, at 24 h.p.i. $n = 6$ for all experiments. The limit of detection is 200 p.f.u. ml⁻¹. **f**, Relative fitness of escape virus at $t = 1$ d.p.i. when competing with ancestral virus on CRISPR-resistant monocultures or polycultures consisting of 48 clones. $n = 6$ for both experiments. **g**, For each diversity treatment shown in Figs 2–4 we examined the spacer content of 192 randomly isolated clones at both $t = 0$ and $t = 3$ d.p.i. (384 clones in total per diversity treatment). The change in the total number of spacers between $t = 0$ and $t = 3$ d.p.i. was calculated independently for each replicate experiment (number of biological replicates as indicated in legend of Fig. 2) and divided by the number of clones that were examined. The graph indicates the average across the replicates of the change in spacer content per clone. **h**, **i**, Titres (in plaque-forming units per millilitre) over time of virus carrying an anti-CRISPR gene (DMS3vir+*acrF1*) following infection of a bacterial population consisting of an equal mixture of a surface mutant and a monoculture with CRISPR-mediated immunity (**h**; $n = 48$) or a 48-clone polyculture with CRISPR-mediated immunity (**i**; $n = 6$). Each clone is equally represented in each treatment. Each line indicates a biological replicate experiment. The limit of detection is 200 p.f.u. ml⁻¹. **j**, The number of replicate experiments in which the CRISPR immune population went extinct (no detectable white colonies) at 1 and 3 d.p.i. In all cases, n is the number of biological replicates and error bars represent 95% CI.



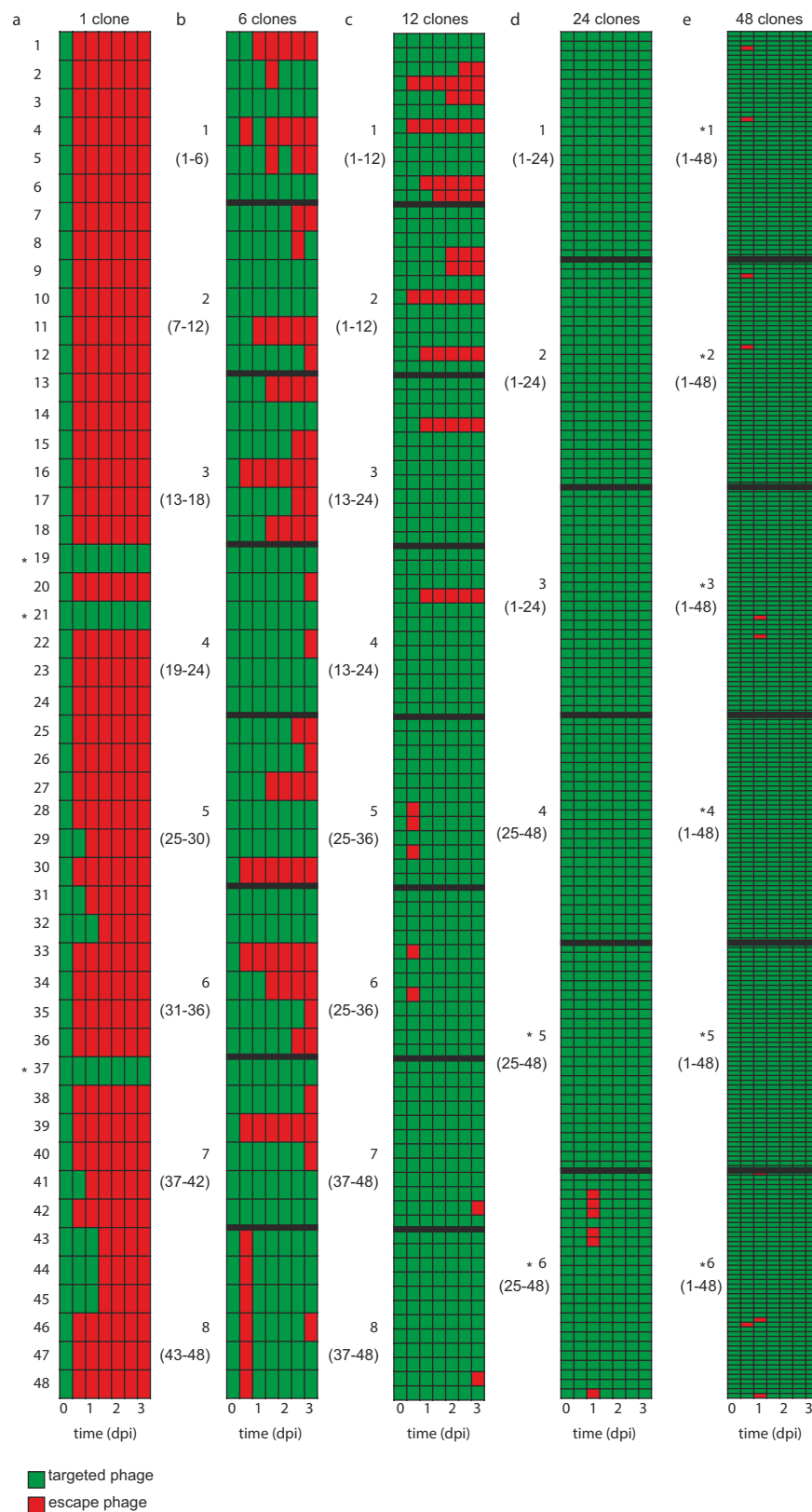
Extended Data Figure 2 | Virus persistence inversely correlates with the level of CRISPR spacer diversity in CRISPR immune populations of *S. thermophilus*. **a, b**, Virus titres (in plaque-forming units per millilitre) over time upon infection of a bacterial population consisting of a monoculture with CRISPR-mediated immunity (**a**; $n = 44$ biological replicates) or 44-clone polycultures with CRISPR-mediated immunity (**b**; $n = 28$ biological replicates). Each clone is equally represented in each

treatment. Each line indicates a biological replicate experiment. The limit of detection is $200 \text{ p.f.u. ml}^{-1}$. **c**, Absorbance at 600 nm of monocultures and polycultures at 16 and 40 h.p.i. Error bars, 95% CI. **d**, Emergence of virus mutants that overcome CRISPR-mediated immunity after 0, 16, 24, 40 and 48 h.p.i. Green indicates no escape virus. Red indicates emergence of escape virus. Escape virus emerged in none of the 28 biological replicates of the polyculture experiments.



Extended Data Figure 3 | Sensitive bacteria are unable to invade bacterial populations with CRISPR-mediated immunity in the presence of virus. a–e, Virus titres (in plaque-forming units per millilitre) over time upon infection of a bacterial population consisting of an equal mixture of a sensitive CRISPR KO clone and a monoculture with CRISPR-mediated immunity (**a**; $n = 48$), or polycultures with CRISPR-mediated immunity consisting of 6 clones (**b**; $n = 8$), 12 clones (**c**; $n = 8$), 24 clones (**d**; $n = 6$) and 48 clones (**e**; $n = 6$). The number of replicates is chosen such that all clones are equally represented in each treatment. Each line indicates a biological replicate experiment. The limit of detection is $200 \text{ p.f.u. ml}^{-1}$.

f, Relative fitness of CRISPR populations at 3 d.p.i. during the competitions with the sensitive CRISPR KO described in **a–e**. Relative fitness of CRISPR populations decreases with increasing spacer diversity due to the rapid virus extinction, which benefits sensitive bacteria, but is higher than 1 in all cases. Error bars, 95% CI. **g**, Relative fitness of monoculture (single spacer; low diversity, $n = 48$) and polyculture (48 spacers; high diversity, $n = 6$) at 3 d.p.i. when competing with the CRISPR KO strain in the absence of virus. Error bars, 95% CI. In all cases n is the number of biological replicates.



Extended Data Figure 4 | Evolution of virus infectivity is constrained by CRISPR diversity when CRISPR immune populations compete with sensitive CRISPR KO bacteria. Emergence of virus mutants that overcome CRISPR-mediated immunity during the experiment shown in Extended Data Fig. 3. Each column in a table represents a time point (0, 16, 24, 40, 48, 64 and 72 h.p.i., as indicated below the table (in days post-infection)) where virus was isolated. Green indicates no escape virus. Red indicates emergence of escape virus. Panels a–e correspond to each of

the experiments shown in Extended Data Fig. 3a–e. Bold numbers indicate each of the individual biological replicates, as detailed in the legend of Extended Data Fig. 3. In b–e, individual replicates are separated by bold lines. Numbers between parentheses indicate the identity of clones that are present in a population with CRISPR-mediated immunity. Asterisks indicate replicate experiments where virus went extinct during the experiment.

Extended Data Table 1 | Tukey's honest significant difference test of all pairwise comparisons of the data in Fig. 3

Comparison		Difference	Std Error	t Ratio	Prob> t	Lower 95%	Upper 95%
1	6	-1.12680	0.2141986	-5.26	<.0001*	-1.72637	-0.52724
1	12	-1.40303	0.2141986	-6.55	<.0001*	-2.00259	-0.80346
1	24	-1.72790	0.2428783	-7.11	<.0001*	-2.40775	-1.04806
1	48	-2.35252	0.2428783	-9.69	<.0001*	-3.03236	-1.67267
6	12	-0.27622	0.2804518	-0.98	0.8612	-1.06124	0.50879
6	24	-0.60110	0.3029225	-1.98	0.2842	-1.44901	0.24682
6	48	-1.22571	0.3029225	-4.05	0.0012*	-2.07363	-0.37780
12	24	-0.32488	0.3029225	-1.07	0.8200	-1.17279	0.52304
12	48	-0.94949	0.3029225	-3.13	0.0205*	-1.79741	-0.10158
24	48	-0.62462	0.3238378	-1.93	0.3119	-1.53108	0.28184

1, Monoculture; 6, 6-clone polyculture; 12, 12-clone polyculture; 24, 24-clone polyculture; 48, 48-clone polyculture.

Metabolic maintenance of cell asymmetry following division in activated T lymphocytes

Katherine C. Verbist¹, Cliff S. Guy¹, Sandra Milasta¹, Swantje Liedmann¹, Marcin M. Kamiński¹, Ruoning Wang² & Douglas R. Green¹

Asymmetric cell division, the partitioning of cellular components in response to polarizing cues during mitosis, has roles in differentiation and development¹. It is important for the self-renewal of fertilized zygotes in *Caenorhabditis elegans* and neuroblasts in *Drosophila*, and in the development of mammalian nervous and digestive systems¹. T lymphocytes, upon activation by antigen-presenting cells (APCs), can undergo asymmetric cell division, wherein the daughter cell proximal to the APC is more likely to differentiate into an effector-like T cell and the distal daughter is more likely to differentiate into a memory-like T cell². Upon activation and before cell division, expression of the transcription factor c-Myc drives metabolic reprogramming, necessary for the subsequent proliferative burst³. Here we find that during the first division of an activated T cell in mice, c-Myc can sort asymmetrically. Asymmetric distribution of amino acid transporters, amino acid content, and activity of mammalian target of rapamycin complex 1 (mTORC1) is correlated with c-Myc expression, and both amino acids and mTORC1 activity sustain the differences in c-Myc expression in one daughter cell compared to the other. Asymmetric c-Myc levels in daughter T cells affect proliferation, metabolism, and differentiation, and these effects are altered by experimental manipulation of mTORC1 activity or c-Myc expression. Therefore, metabolic signalling pathways cooperate with transcription programs to maintain differential cell fates following asymmetric T-cell division.

To visualize c-Myc levels in activated T cells, we isolated T cells from c-Myc-GFP fusion knock-in (c-Myc-GFP) mice⁴ and activated them *in vitro* with anti-CD3 antibodies, anti-CD28 antibodies, and ICAM². As T cells completed the first division (indicated by dilution of CellTrace Violet), the c-Myc-GFP signal was brightest in cells that expressed higher levels of CD8, a marker of asymmetric cell division² (Fig. 1a, Extended Data Fig. 1a). This difference between CD8^{high} and CD8^{low} cells dissipated in subsequent divisions, as did the difference in c-Myc expression (Fig. 1a, Extended Data Fig. 1a). The asymmetric segregation of c-Myc was also assessed by confocal microscopy 36 h after activation. The largest numbers of first-division T cells were recovered at this time point (Extended Data Fig. 1b). Again, an asymmetric inheritance of c-Myc-GFP was consistently observed in daughter T cells that expressed higher levels of CD8 (Fig. 1b, c, Extended Data Fig. 1c, Supplementary Videos 1–3).

The first division of a T cell takes place in contact with an APC, so we next determined whether c-Myc preferentially localizes to the proximal or distal daughter². To this end, c-Myc-GFP, OT-I transgenic (OT-I Tg) T cells were activated by bone-marrow-derived dendritic cells (BMDCs) pulsed with SIINFEKL (the OT-I T cell cognate antigen). By analysing conjoined daughter cells via microscopy, we observed that c-Myc was asymmetrically inherited by the proximal CD8^{high} daughter cell (Fig. 1d, e, Supplementary Videos 4–6, Extended Data Fig. 1c). We then analysed expression of several markers of

asymmetric cell division¹. As expected, Numb and Scribble were enriched in proximal daughters, along with c-Myc-GFP (Fig. 1f, g), whereas PKC ζ was enriched in distal daughters (Fig. 1h, Extended Data Fig. 2).

When activated *in vitro*, c-Myc asymmetrically segregated in first-division daughter cells in a substantial proportion of T cells (Extended Data Fig. 1d). To determine whether this also occurred *in vivo*, carboxyfluorescein succinimidyl ester (CFSE)-labelled OT-I T cells were transferred into congenic hosts. Recipients were then infected with *Listeria monocytogenes* SIINFEKL. On day 3 after infection, confocal analysis of CFSE⁺ conjoined T-cell pairs revealed that c-Myc staining was significantly higher in one T cell compared to the other; therefore, the asymmetric assortment of c-Myc probably occurs *in vivo* in response to infection (Fig. 1i, j).

Real-time analysis of GFP expression during mitosis revealed that the signal was diffuse throughout the cell until after division. The signal then increased in one daughter cell, establishing an asymmetric distribution (Fig. 2a and Supplementary Video 7). In fixed T cells, we observed the GFP signal was diffuse from prophase through anaphase, and only upon cytokinesis and re-formation of the nuclear envelope were c-Myc levels distinguishable in the daughter cells (Fig. 2b, Extended Data Fig. 3). It is therefore likely that c-Myc is differentially regulated in the two daughters by asymmetrically inherited upstream signalling proteins, rather than being polarized itself.

To determine if differences in the levels of c-Myc following the first division are relevant to c-Myc function, we examined undivided and first-division T cells sorted into c-Myc^{high} and c-Myc^{low} populations for expression of several metabolic genes that were previously found to be controlled by c-Myc³. We found similar differences in expression of most of these genes influenced by c-Myc-GFP in both undivided and first-division cells (Extended Data Fig. 4a–c). Therefore, the difference between c-Myc^{low} and c-Myc^{high} upon asymmetric cell division is relevant for expression of c-Myc target genes.

We assessed several activation markers on c-Myc^{high} and c-Myc^{low} T cells before and after the first division. CD44 expression was comparable among all populations, and both c-Myc^{high} and c-Myc^{low} populations exhibited increased expression of CD69 compared with undivided cells. Although all activated cells also displayed increased CD25 and CD98, c-Myc^{high} T cells displayed elevated levels of both (Fig. 2c), as previously described for CD25 (ref. 2). Interleukin (IL)-2 can drive the expression of c-Myc⁵, but neither transient inhibition of IL-2 receptor signalling with janus kinase (JAK) inhibitors (Extended Data Fig. 5a) nor increased IL-2 (Extended Data Fig. 5b) influenced the asymmetric assortment of c-Myc in T cells following the first division.

We then further examined the CD98 asymmetry. Microscopy confirmed that surface expression of CD98 was elevated on the proximal daughter T cell and correlated with c-Myc-GFP expression (Fig. 2d, e, Extended Data Fig. 5c, d). CD98 is a heterodimer composed

¹Department of Immunology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. ²Center for Childhood Cancer and Blood Disease, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA.

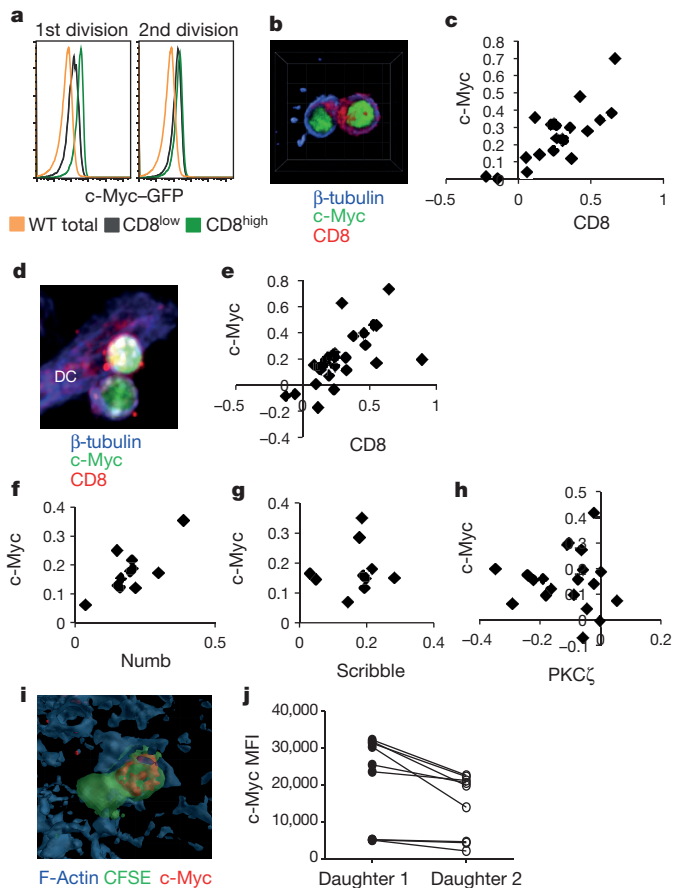


Figure 1 | c-Myc asymmetrically segregates to the proximal daughter cell in activated CD8 T lymphocytes. **a**, Mean fluorescent intensities (MFI) of c-Myc-GFP in negative (wild-type (WT)), CD8^{low}, and CD8^{high} cells in the first (left) and second (right) divisions. Representative of four independent experiments. **b**, Representative image of conjoined daughter c-Myc-GFP CD8 T cells (antibody-coated plates) fixed and stained for β -tubulin (blue) and CD8 (red). **c**, Quantification of asymmetry based on fluorescent intensities of CD8 (difference/total; x axis) and c-Myc-GFP (difference/total; y axis); 88.9% concordance of markers in the same daughter cell ($P = 0.0004$, two-tailed binomial test; $R^2 = 0.6159$, $P < 0.0001$ linear regression). Compiled from four independent experiments; each point represents a conjoined daughter pair. **d**, **e**, Representative image and quantification of asymmetry of conjoined daughter OT-I cells co-cultured with BMDCs; 86.2% both bright in proximal daughter ($\chi^2 = 58.31$, degrees of freedom (DF) = 3, $P < 0.0001$, chi-square goodness of fit; $R^2 = 0.3939$, $P = 0.0003$, linear regression). **f**–**h**, Quantifications of asymmetry (difference/total) of asymmetric cell division markers: Numb 100% both bright in proximal daughter ($\chi^2 = 33$, DF = 3, $P < 0.0001$, chi-square goodness of fit; $R^2 = 0.5738$, $P = 0.0069$, linear regression) (**f**); Scribble 100% both bright in proximal daughter ($\chi^2 = 30$, DF = 3, $P < 0.0001$, chi-square goodness of fit; $R^2 = 0.01436$, $P = 0.7416$, linear regression) (**g**); and PKC ζ 85% bright in distal daughter, c-Myc bright in proximal daughter ($\chi^2 = 26.8$, DF = 3, $P < 0.0001$, chi-square goodness of fit; $R^2 = 0.0009705$, $P = 0.8963$, linear regression) (**h**). **i**, **j**, Representative image and quantification of c-Myc fluorescence of dividing OT-I T cells in spleen sections of animals infected with *M. listeria* OVA, $P = 0.0077$; paired Student's *t*-test. Compiled from two independent experiments. Original magnification, $\times 100$ (**b**, **d**, **i**).

of SLC3A2 and SLC7A5. Analyses of these components in undivided cells revealed that SLC3A2 protein is polarized to the region of the T cell near the contact site during activation, before cell division (Extended Data Fig. 5e, f), which probably contributes to its asymmetry. Similarly, expression of the neutral amino acid transporter SLC1A5 was higher on c-Myc^{high} daughter T cells (Extended Data Fig. 6a). In contrast, no difference in expression of the aspartate and glutamate transporter SLC1A3 was observed (Extended Data Fig. 6b). Evaluation

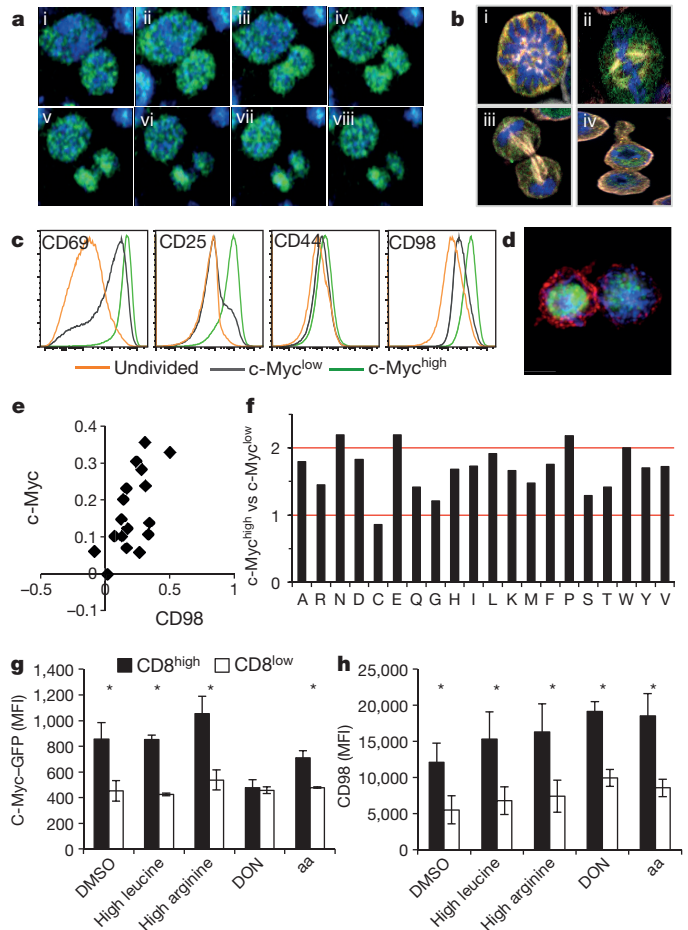


Figure 2 | Amino acid metabolism is necessary for the maintenance of c-Myc asymmetry in activated CD8 T cells. **a**, Time-lapse of dividing c-Myc-GFP OT-I T cells co-cultured with BMDCs at 4 min intervals (i–viii). **b**, Fixed T cells (antibody-coated plates) stained with Hoechst 33258 (blue) and anti- β -tubulin (white) to identify the stages of mitosis: prophase (i), metaphase (ii), anaphase (iii), telophase/cytokinesis (iv). **c**, MFI of indicated activation markers for activated undivided T cells (gold), first-division c-Myc^{low} T cells (grey), or first-division c-Myc^{high} T cells (green) (antibody-coated plates). Representative of four independent experiments. **d**, **e**, Representative image and quantification of fluorescent intensity (difference/total) of CD98 (red) in T cells co-cultured with BMDCs; 88.2% both bright in proximal daughter ($\chi^2 = 36.41$, DF = 3, $P < 0.0001$, chi-square goodness of fit; $R^2 = 0.3886$, $P = 0.0075$, linear regression). **f**, Ratio of amino acids in c-Myc^{high} versus c-Myc^{low} first-division CD8 T cells. Values normalized in terms of raw area counts. **g**, **h**, MFI of c-Myc-GFP (**g**) or CD98 (**h**) in first-division CD8^{high} and CD8^{low} (35 h on antibody-coated plates) and indicated treatment (DMSO (dimethyl sulfoxide control), excess leucine or arginine in media, or 6-diazo-5-oxo-L-norleucine (DON)) for 45 min (aa, amino acid starvation; four hours). Mean \pm s.d. ($n = 3$ mice per group). $*P < 0.05$. Representative of two independent experiments. Original magnification, $\times 100$ (**a**, **b**, **d**).

of metabolites revealed that levels of several, but not all, amino acids were at least 1.5-fold higher in first-division c-Myc^{high} versus c-Myc^{low} T cells (Fig. 2f). We therefore examined the effects of amino acid availability on the asymmetric expression of c-Myc. Although high leucine or glutamine had no effect on c-Myc expression in first-division T cells, transient depletion of all amino acids greatly diminished c-Myc asymmetry (Fig. 2g). Asymmetric distribution of c-Myc-GFP was sensitive to transient inhibition of glutaminolysis by the glutamine analogue 6-diazo-5-oxo-L-norleucine (Fig. 2g). Expression levels of CD98 were unchanged in these conditions (Fig. 2h). Thus, amino acid availability as well as glutaminolysis appears to sustain high c-Myc levels following asymmetric cell division.

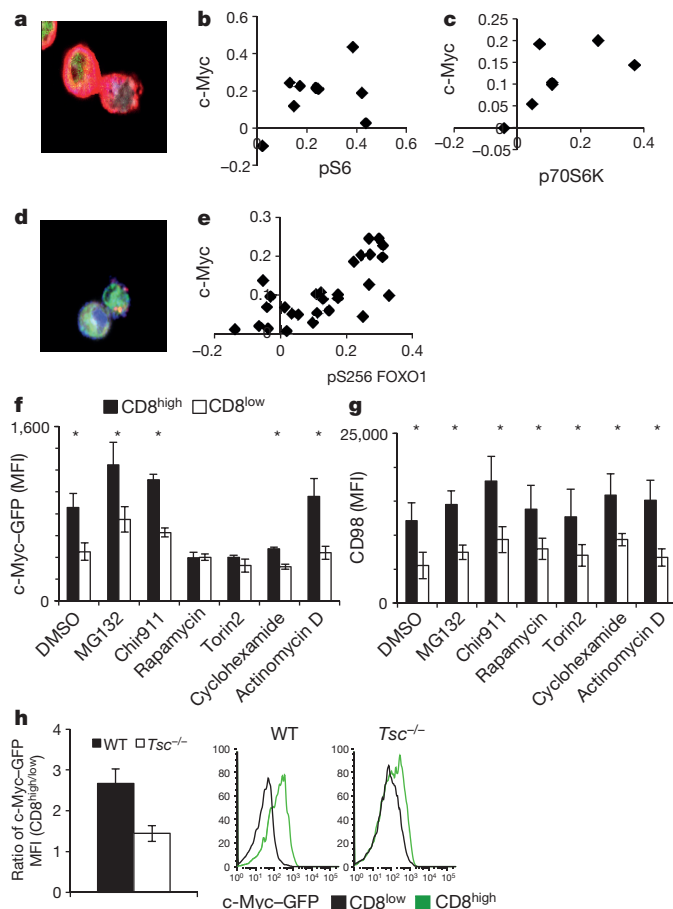


Figure 3 | mTORC1 activity is required for the maintenance of c-Myc asymmetry in activated CD8 T cells. **a–c**, Representative image and quantification of asymmetry based on fluorescence intensity (difference/total) of pS6 (red); 88.9% both bright in proximal daughter ($\chi^2 = 19.89$, DF = 3, $P = 0.0002$, chi-square goodness of fit; $R^2 = 0.1388$, $P = 0.1388$, linear regression) (**a**, **b**) and p70S6K (red); 85.7% both bright in proximal daughter ($\chi^2 = 14.14$, DF = 3, $P = 0.0027$, chi-square goodness of fit; $R^2 = 0.3526$, $P = 0.1598$, linear regression) (**c**) in T cells co-cultured with BMDCs. **d**, **e**, Representative image and quantification of asymmetry based on fluorescence intensity (difference/total) of pFOXO1 (pS256) (red) in CD8 T cells (antibody-coated plates); 75.8% concordance of markers ($P = 0.0081$, two-tailed binomial test; $R^2 = 0.5153$, $P < 0.0001$, linear regression). Original magnification, $\times 100$ (**a**, **d**). **f**, **g**, MFI of first division c-Myc-GFP (**f**) or CD98 (**g**) in CD8^{high} and CD8^{low} after 35 h activation on antibody-coated plates and the indicated agent for 45 min before flow cytometric analysis. * $P < 0.05$, unpaired Student's *t*-test. **h**, CellTrace-Violet-labelled T cells from lymphoid tissues of CD4-Cre⁻ (WT) or CD4-Cre⁺ (*Tsc*^{-/-}) *Tsc*^{flx/flx} animals 36 h after activation. The ratio of MFI of c-Myc-GFP in first-division CD8^{high}/CD8^{low} T cells is plotted (left), and representative flow plots are depicted (right).

One consequence of amino acid availability is activation of the mTORC1 complex, which is also activated by glutaminolysis^{6,7}. Both phospho (p)-mTOR distribution (Extended Data Fig. 7a, b) and mTORC1 activity⁸, assessed by pS6 (Fig. 3a, b) or p70S6K (Fig. 3c), associated with c-Myc^{high} daughter T cells (Extended Data Fig. 8a, b). Reciprocally, inhibition of c-Myc expression by treatment with the inhibitor JQ1 resulted in reduced mTORC1 signalling (Extended Data Fig. 7c), indicating a positive feedback mechanism. Another indirect target, phosphorylated when mTORC1 is active, is the transcription factor FOXO1⁹. FOXO1 is important for maintaining quiescence in T cells¹⁰ and negatively regulates c-Myc target genes in its non-phosphorylated form¹¹. Phosphorylated FOXO1 (pFOXO1) was also correlated with c-Myc-GFP expression (Fig. 3d, e, Extended Data Fig. 8e).

We then acutely inhibited mTORC1 activity following the first division. Inhibition of mTORC1 activity with either rapamycin or Torin2 reduced c-Myc expression in CD8^{high} T cells compared to that of CD8^{low} cells (Fig. 3f), but did not influence CD98 asymmetry (Fig. 3g). To determine if elevation of mTORC1 activity affects c-Myc asymmetry, CD4-Cre, *Tsc1*^{flx/flx} mice, which exhibit hyperactive mTORC1 signalling¹², were crossed with c-Myc-GFP mice. Unlike those from wild-type littermates, *Tsc1*^{-/-} CD8^{high} and CD8^{low} T cells in the first division had similarly high expression of c-Myc-GFP (Fig. 3h). Therefore, asymmetric mTORC1 activity sustains the asymmetric assortment of c-Myc following T-cell division.

Various signals determine when and how much c-Myc is expressed¹³. Although phosphorylation of c-Myc by glycogen synthase kinase 3 (GSK3) promotes its proteasomal degradation¹⁴, GSK3 inhibition had no effect on c-Myc asymmetry (Fig. 3f). Further, although acute inhibition of the proteasome with MG132 elevated c-Myc expression in both CD8^{high} and CD8^{low} daughter cells, asymmetry was unaffected. Strikingly, acute inhibition of translation by cyclohexamide attenuated c-Myc expression and asymmetry, whereas acute inhibition of transcription by actinomycin D had no effect (Fig. 3f). None of the short-term treatments tested had any effect on the expression or asymmetric distribution of CD98 (Fig. 3f). Therefore, it is likely that mTORC1 activity regulates c-Myc at the level of translation, although a role for transcription over a longer timeframe could not be assessed. Consistently, we observed that mRNA for c-Myc and both components of CD98 (SLC3A2 and SLC7A5), and several other metabolic genes (which are c-Myc targets¹⁵) were enriched in first-division c-Myc^{high} T cells compared to c-Myc^{low} T cells (Extended Data Fig. 4a, c).

T-cell proliferation depends on c-Myc⁸. We therefore examined if the relative differences in c-Myc levels in the first division influence subsequent proliferation. At 35 h after activation, CellTrace-Violet-labelled activated T cells were pulsed with BrdU for 1 h. The cell cycle status of c-Myc^{high} and c-Myc^{low} T cells that had divided once was examined, and nearly ten times the frequencies of c-Myc^{high} T cells were cycling (in S or G2/M phases) compared with c-Myc^{low} T cells (Fig. 4a). Additionally, sorted c-Myc^{high} T cells proliferated much more rapidly than c-Myc^{low} T cells (Fig. 4b, c) when placed back in culture for 48 h. In contrast, sorted CD8^{high} and CD8^{low} first-division TSC1-null T cells, which display similar c-Myc levels (Fig. 3h), showed equivalent proliferation (Fig. 4c).

As c-Myc regulates metabolic reprogramming during T-cell activation^{3,16,17}, we tested whether metabolic differences exist in first-division T cells expressing high or low levels of c-Myc. We found that c-Myc^{high} cells possessed diminished levels of glucose and glycolytic intermediates, but elevated levels of downstream metabolites (3-phosphoglycerate and phosphoenolpyruvate) and end products pyruvate and lactate, which may reflect enhanced glycolysis in c-Myc^{high} cells. Elevated levels of tricarboxylic acid cycle intermediates such as α -ketoglutarate, fumarate, and malate in c-Myc^{high} T cells may reflect enhanced glutaminolysis (Extended Data Fig. 9). Additionally, pentose phosphate pathway metabolites including 6-phosphogluconate and sedoheptulose 7-phosphate accumulated in c-Myc^{high} cells. We then examined the extracellular acidification rate in c-Myc^{high} and c-Myc^{low} T cells under conditions of glycolytic stress^{18,19} and found that c-Myc^{high} T cells tended to be more glycolytic (Fig. 4e, Extended Data Fig. 10a). Even during basal respiration with glucose, c-Myc^{high} T cells exhibited higher extracellular acidification rate with no changes in oxygen consumption rate (Extended Data Fig. 10b). Similarly, metabolic flux assays tracking radiolabelled substrates revealed increased glycolysis and glutamine uptake in c-Myc^{high} T cells (Fig. 4d). Although it has been suggested that mitochondria sort asymmetrically²⁰, we did not detect any differences in mitochondrial mass or DNA in c-Myc^{low} compared with c-Myc^{high} first-division T cells (Extended Data Fig. 10c, d), but this does not exclude the possibility that 'older' mitochondria may sort preferentially into one daughter²⁰. However, we suggest that the differences in

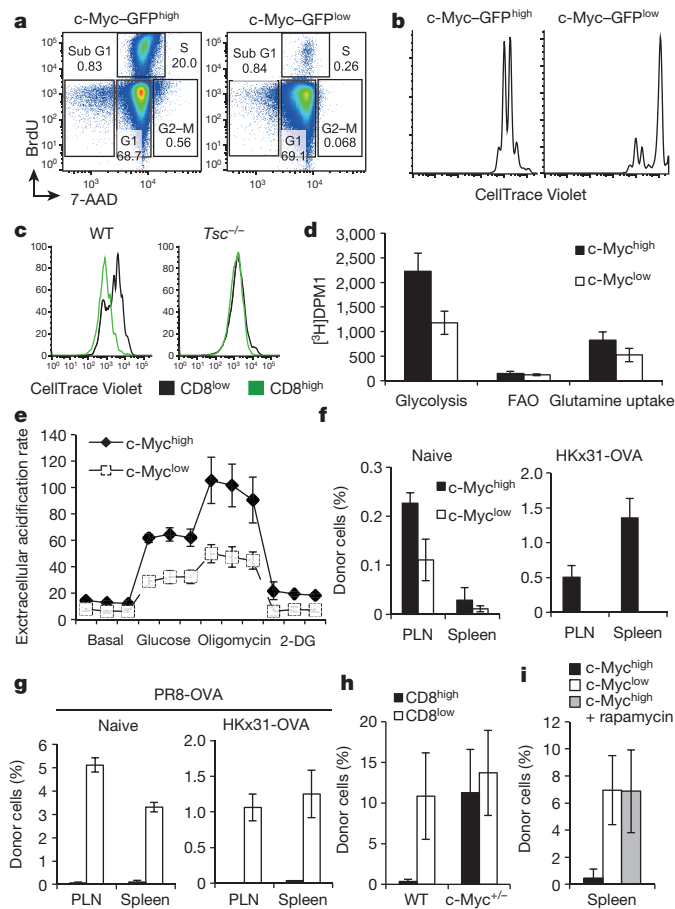


Figure 4 | T cell proliferation, metabolism, and fate are directed by differential expression of c-Myc protein. **a**, Representative flow plots for first-division c-Myc-GFP^{high} (left panel) and c-Myc-GFP^{low} (right panel) cells (antibody-coated plates and 1 h BrdU staining). **b**, CellTrace Violet dilution of antibody-activated, sorted first-division c-Myc-GFP^{high} and c-Myc-GFP^{low} T cells after culture for a further 38 hours. **c**, Populations in Fig. 3h, sorted and cultured for a further 48 h; CellTrace Violet dilutions of CD8^{high} cells and CD8^{low} cells shown. **d**, Sorted first-division CD8 T cell with high or low c-Myc-GFP expression were re-plated with [³-H]glucose, [9,10-³H]palmitic acid, or [³H]glutamine. Mean \pm s.d., disintegrations per minute of triplicates ($n = 3$ mice per group). **e**, Extracellular acidification rate of sorted first-division c-Myc^{high} and c-Myc^{low} T cells during exposure to indicated compounds. Mean \pm s.d. for three replicates, pooled from two mice. Representative of three independent experiments (Extended Data Fig. 10). **f**, Frequency of donor cells two weeks after transfer of 5×10^5 first-division c-Myc-GFP^{high} or c-Myc-GFP^{low} OT-I cells into naive recipients infected, or not, with HKx31 influenza expressing the SIINFEKL peptide. **g**, Frequencies of OT-I donor cells 9 days after challenge of the animals in **f** with PR8 influenza expressing SIINFEKL peptide (after an additional 2 weeks rest), mean \pm s.d. ($n = 3$ mice per group, representative of two independent experiments). **h**, Frequencies of donor cells from c-Myc WT ($P = 0.0059$, Student's *t*-test) or c-Myc^{+/-} mice ($P = 0.4040$, Student's *t*-test) 9 days after influenza-OVA challenge (mean \pm s.d., $n \geq 4$ mice per group). **i**, Frequencies of donor cells from recipients treated with PBS ($P = 0.0128$, Student's *t*-test) or rapamycin ($P = 0.9726$, Student's *t*-test) 9 days after influenza-OVA challenge (mean \pm s.d., $n \geq 3$ mice per group).

metabolism are the results of asymmetric distribution of amino acid transporters, mTORC1 activity, and c-Myc.

We next explored functional consequences of the asymmetric distribution of c-Myc *in vivo*. First-division T cells were sorted on the basis of c-Myc-GFP expression and transferred into separate, congenic recipients. Host mice were naive or were infected with influenza A virus (strain HKx31 (H3N2)) expressing SIINFEKL. Two weeks later, a greater frequency of donor OT-I cells was recovered from the

lymphoid organs of mice that had received c-Myc^{high} T cells (Fig. 4f). After an additional two weeks, recipient animals were challenged with a heterosubtypic virus (restricting responses to T cells) expressing SIINFEKL. Nine days later, an inverse pattern to pre-challenge results was observed, with increased frequencies of cells derived from the c-Myc^{low} T cells (Fig. 4g). Thus, the asymmetric distribution of c-Myc in the first division was directly correlated with the eventual fate of these cells, with high levels of c-Myc conferring increased proliferation in the primary response and low levels of c-Myc conferring decreased proliferation in the primary response (but increased persistence and secondary response). Importantly, when T cells were wild type or heterozygous for c-Myc (exhibiting reduced c-Myc expression), sorted CD8^{high} and CD8^{low} T cells from c-myc^{+/-} T cells contributed equally well to the recall response, whereas wild-type CD8^{high} T cells did not (Fig. 4h). *In vivo* inhibition of mTORC1 in recipient animals with rapamycin also restored the ability of CD8^{high} T cells to contribute to a secondary response (Fig. 4i), underscoring the relationship between mTORC1 signalling and c-Myc expression in regulating cell fate. These findings corroborate studies that show distinct fates of T cell daughters that have undergone asymmetric cell division, with the proximal and distal daughters displaying effector-like and memory-like gene expression and function²¹, but suggest that these fates remain amenable to regulation (for example, of mTORC1 or c-Myc). Although the fates of asymmetrically divided T cells are distinct, we do not suggest this is necessarily the only source of memory T-cell differentiation. Other studies have shown that memory cells have expressed proteins such as granzyme B and interferon (IFN) γ ²²⁻²⁴, and are thus probably products of effector cells^{25,26}. Our data support a model in which T cells proceed through an activation phase, but asymmetric division results in a segregation of activating signals and amino acid transporters that result in divergent mTORC1 activity and c-Myc expression that ultimately make a daughter cell more or less likely to become terminally differentiated.

Our findings suggest that the interplay of amino acid transport and glutaminolysis, mTORC1 activity, and c-Myc expression sustains the fates of T cells that have undergone asymmetric cell division. c-Myc drives the expression of the amino acid transporters CD98 and SLC1A5^{3,27}, and sustains mTORC1 activity during early activation²⁸. Conversely, amino acids and glutaminolysis sustain mTORC1 activity, which sustains c-Myc translation^{7,29}, with possible effects on c-Myc through inhibition of FOXO1¹¹. During activation, some amino acid transporters sort asymmetrically, and a positive feedback arises, causing asymmetry in mTORC1 activity and c-Myc expression, which is then sustained by c-Myc function (or lack thereof) in the daughter cells. Notably, c-Myc was among the top candidates likely to drive divergence in gene expression between the daughters of first division in activated T cells²¹. Consistently, T cells lacking CD98 or treated with inhibitors of mTORC1³⁰ display poor primary responses but generate memory T cells, whereas T cells lacking TSC1 (and elevated mTORC1 activity)¹² or lacking FOXO1, display robust primary responses but poor memory responses¹⁰. As c-Myc is also clearly consequential for oncogenesis, it is also possible that lymphomas of mature T cells might be influenced by symmetric cell division. A deeper understanding of asymmetric T-cell division as it applies to mTORC1 and c-Myc, may help to explain these observations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 April 2015; accepted 23 February 2016.

Published online 11 April 2016; corrected online 20 April 2016

(see full-text HTML version for details).

- Pham, K., Sacirbegovic, F. & Russell, S. M. Polarised cells, polarised views: asymmetric cell division in hematopoietic cells. *Front. Immunol.* **5**, (2014).
- Chang, J. T. *et al.* Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* **315**, 1687–1691 (2007).
- Wang, R. *et al.* The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35**, 871–882 (2011).

4. Huang, C. Y., Bredemeyer, A. L., Walker, L. M., Bassing, C. H. & Sleckman, B. P. Dynamic regulation of c-Myc proto-oncogene expression during lymphocyte development revealed by a GFP-c-Myc knock-in mouse. *Eur. J. Immunol.* **38**, 342–349 (2008).
5. Colombetti, S., Basso, V., Mueller, D. L. & Mondino, A. Prolonged TCR/CD28 engagement drives IL-2-independent T cell clonal expansion through signaling mediated by the mammalian target of rapamycin. *J. Immunol.* **176**, 2730–2738 (2006).
6. Sancak, Y. *et al.* The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **320**, 1496–1501 (2008).
7. Grohmann, U. & Bronte, V. Control of immune response by amino acid metabolism. *Immunol. Rev.* **236**, 243–264 (2010).
8. Choo, A. Y., Yoon, S. O., Kim, S. G., Roux, P. P. & Blenis, J. Rapamycin differentially inhibits S6Ks and 4E-BP1 to mediate cell-type-specific repression of mRNA translation. *Proc. Natl Acad. Sci. USA* **105**, 17414–17419 (2008).
9. Hsu, P. P. *et al.* The mTOR-Regulated Phosphoproteome Reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332**, 1317–1322 (2011).
10. Ouyang, W., Beckett, O., Flavell, R. A. & Li, M. O. An essential role of the Forkhead-box transcription factor Foxo1 in control of T cell homeostasis and tolerance. *Immunity* **30**, 358–371 (2009).
11. Peck, B., Ferber, E. C. & Schulze, A. Antagonism between FOXO and Myc regulates cellular powerhouse. *Front. Oncol.* **3**, (2013).
12. Yang, K., Neale, G., Green, D. R., He, W. & Chi, H. The tumor suppressor Tsc1 enforces quiescence of naive T cells to promote immune homeostasis and function. *Nature Immunol.* **12**, 888–897 (2011).
13. Vervoorts, J., Luscher-Firzlaff, J. & Luscher, B. The ins and outs of MYC regulation by posttranslational mechanisms. *J. Biol. Chem.* **281**, 34725–34729 (2006).
14. Sears, R. C. The life cycle of C-myc: from synthesis to degradation. *Cell Cycle* **3**, 1133–1137 (2004).
15. Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA* **105**, 18782–18787 (2008).
16. Wang, R. & Green, D. R. Metabolic checkpoints in activated T cells. *Nature Immunol.* **13**, 907–915 (2012).
17. Marelli-Berg, F. M., Fu, H. & Mauro, C. Molecular mechanisms of metabolic reprogramming in proliferating cells: implications for T-cell-mediated immunity. *Immunology* **136**, 363–369 (2012).
18. van der Windt, G. J. *et al.* Mitochondrial respiratory capacity is a critical regulator of CD8⁺ T cell memory development. *Immunity* **36**, 68–78 (2012).
19. van der Windt, G. J. & Pearce, E. L. Metabolic switching and fuel choice during T-cell differentiation and memory development. *Immunol. Rev.* **249**, 27–42 (2012).
20. Katajisto, P. *et al.* Stem cells. Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness. *Science* **348**, 340–343 (2015).
21. Arsenio, J. *et al.* Early specification of CD8⁺ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nature Immunol.* **15**, 365–372 (2014).
22. Stemmerger, C. *et al.* A single naive CD8⁺ T cell precursor can develop into diverse effector and memory subsets. *Immunity* **27**, 985–997 (2007).
23. Schepers, K. *et al.* Dissecting T cell lineage relationships by cellular barcoding. *J. Exp. Med.* **205**, 2309–2318 (2008).
24. Bannard, O., Kraman, M. & Fearon, D. T. Secondary replicative function of CD8⁺ T cells that had developed an effector phenotype. *Science* **323**, 505–509 (2009).
25. Harrington, L. E., Janowski, K. M., Oliver, J. R., Zajac, A. J. & Weaver, C. T. Memory CD4 T cells emerge from effector T-cell progenitors. *Nature* **452**, 356–360 10.1038/nature06672 (2008).
26. Rohr, J. C., Gerlach, C., Kok, L. & Schumacher, T. N. Single cell behavior in T cell differentiation. *Trends Immunol.* **35**, 170–177 (2014).
27. Dang, C. V., Le, A. & Gao, P. MYC-induced cancer cell energy metabolism and therapeutic opportunities. *Clin. Cancer Res.* **15**, 6479–6483 (2009).
28. Gao, P. *et al.* c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–765 (2009).
29. Liu, J. & Levens, D. Making myc. *Curr. Top. Microbiol. Immunol.* **302**, 1–32 (2006).
30. Araki, K. *et al.* mTOR regulates memory CD8 T-cell differentiation. *Nature* **460**, 108–112 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Cross, G. Lennon, and P. Ingle for cell sorting, P. Thomas for assistance with influenza infections, H. Chi for help with *Listeria* infections, and I. Pawlikowska for help with statistical analyses. This work was supported by ALSAC and grants from the US National Institutes of Health.

Author Contributions K.C.V. conceived the project, designed and performed most experiments, interpreted results, and co-wrote the manuscript. C.S.G. guided imaging, performed live imaging experiments, and performed experiments with TSC1 flox/flox animals. S.M. provided technical assistance with Seahorse XF24 assays and helped with several *in vivo* studies. S.L. performed some imaging experiments with amino acid transports, performed experiments with high IL-2, and helped with several *in vivo* studies. M.M.K. performed and analysed all gene expression analyses. R.W. provided intellectual contributions. D.R.G. conceived the project, supervised experimental designs, interpreted results, and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R.G. (douglas.green@stjude.org).

METHODS

Animals. c-Myc–GFP fusion knock-in mice were generated and provided by B. Sleckman⁴. OT-I Tg (C57BL/6-Tg (Tcr α Tcr β)1100Mjb/J) were acquired from The Jackson Laboratory (Bar Harbour, Maine), and TSC1 flox (Tsc1^{tm1Djk}/J) mice were a gift from H. Chi (St. Jude Children's Research Hospital). Mice heterozygous for c-Myc were generated from c-Myc flox mice, a gift from F. Alt (Boston Children's Hospital) (see Supplementary Methods). The experiments were not randomized; all animal experiments were conducted with either male or female sex and aged-matched littermate controls (6–16 weeks old). No statistical methods were used to predetermine sample size; rather, sample sizes were selected based on animal availability and experience with variation within the immune system. Experimental analyses were not blinded. St. Jude Institutional Animal Care and Use Committee approved all procedures in accordance with the Guide for the Care and Use of Animals.

T-cell stimulations. For polyclonal T cells, cells were isolated from the spleens and lymph nodes of c-Myc–GFP mice by mechanical passage through a Falcon 70 μ m cell strainer (Thermo Fisher Scientific) and lysis of red blood cells in hypotonic solution. Single-cell suspensions were enriched for T cells using CD45R (B220) microbeads and MACS separation (Miltenyi). Enriched cells were then stimulated on plate-bound anti-CD3 (1 μ g ml⁻¹), anti-CD28 (1 μ g ml⁻¹) (BioXCell), and recombinant human CD54/ICAM (0.5 μ g ml⁻¹) produced in insect cells for 30–36 h¹. OT-I Tg T cells were similarly isolated and enriched to ~96% purity and were either stimulated as above or overlain onto BMDCs previously pulsed with 100 nM of SIINFEKL peptide (60193-1, Anaspec Inc.) at 37°C for 1 h. Co-cultures were incubated at 37°C for 36–40 h. For Extended Data Fig. 5c, OT-I Tg T cells stimulated with CD3, CD28, and ICAM for 2 h in #1.5 polymer coverslips (80426, ibidi) coated. In all stimulations for microscopy experiments, nocodazole (100 ng ml⁻¹) (Sigma-Aldrich) was added to culture for 4 h and washed off before subsequent imaging. All T cells were cultured in RPMI 1640 (Gibco) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS), 2 mM L-glutamine, 0.05 mM 2-mercaptoethanol, 100 U ml⁻¹ penicillin and 100 μ g ml⁻¹ streptomycin at 37°C in 5% CO₂.

T-cell imaging. Antibodies against the following antigens were used: CD8 (13-0081-85; eBioscience), β -tubulin (T8328; Sigma-Aldrich), Numb (ab14140; Abcam), Scribble (ab154067; Abcam), PKC ζ (IMG-90589-2; Imgenex), c-Myc (9402; Cell Signaling), pS6 (5364; Cell Signaling), p70S6K (9234; Cell Signaling), pmTOR (2974; Cell Signaling), pS256 FOXO1 (9461; Cell Signaling), CD98-PE (12-0981-81; eBiosciences), SLC1A5 (ARP42247_T100; AVIVA), and SLC1A3 (ABIN377559; antibodiesonline). Additionally, DAPI (D9542; Sigma-Aldrich) or Hoechst 33258 (H3569; Invitrogen) was used to identify nuclei.

Methodology for imaging of asymmetrically dividing T cells was adapted from ref. 2. T-cell pairs undergoing cytokinesis were identified by dual nuclei and pronounced cytoplasmic cleft by brightfield. Pairs were then confirmed to be connected based on β -tubulin staining. Subsequently, the morphology of the other fluorescence channels was revealed. All imaging was conducted at $\times 100$ magnification. For quantification, the two nascent daughters were delineated via the pattern of tubulin fluorescence to define the border of each daughter cell. Using Slidebook imaging software, delineated regions were converted to masks, and mask statistics of the sum fluorescent intensity of the GFP channel (for c-Myc) and red fluorescent protein (RFP) channel (for indicated antigen) were analysed using the following formulas: (1) (difference in RFP fluorescence intensity in GFP high – low daughter)/(sum of RFP fluorescence intensity in GFP high and low daughter) for antibody-activated T cells; (2) (difference in RFP fluorescence intensity in proximal daughter – distal daughter)/(sum of RFP fluorescence intensity in both daughters).

Flow cytometry. Antibodies from eBiosciences included anti-CD8-APC-eFluor780 (47-0081), anti-CD44-APC (17-0441-81), anti-CD69-PERCPcy5.5 (45-0691), anti-CD98-PE (12-0981-81; eBiosciences), and anti-CD62L-PEcy7 (25-0621). Anti-CD4-BD605NC was acquired from BD Biosciences. Cells were stained for 20 min at 4°C in PBS, 5% BSA, 0.1% NaN₃. Cell proliferation was assessed using CellTrace Violet (C34557; Life Technologies) according to the manufacturer's recommendations. Samples were acquired on a BD LSRII flow cytometer and analysed with Treestar FlowJo software. Cell cycle analysis was performed using BD APC BrdU Flow kit as per the manufacturer's instructions (552598; BD Biosciences). Cells were sorted on a MoFlow (Beckman-Coulter) or Reflection (i-Cyt).

Compounds. Chemical inhibitors used included 1 μ M MG132 (S2619; Selleckchem), 1.3 μ M Chir911, 1 μ M rapamycin (S1039, Selleckchem), 1 μ M Torin (S2817; Selleckchem), 1 μ g ml⁻¹ DON (D2141, Sigma-Aldrich), 1 μ g ml⁻¹ α -ketoglutarate (K2000; Sigma-Aldrich), 0.1 μ M actinomycin D (114666; Calbiochem), 100 μ g ml⁻¹ cyclohexamide (C7698; Sigma-Aldrich), 1 μ M 2-deoxy-D-glucose (2DG) (D6134; Sigma-Aldrich), and 1 or 5 μ M JQ1

(A1910; Apexbio). Cytokines included recombinant murine IL-7 (5 ng ml⁻¹) (217-17; Peprotech), recombinant murine IL-2 (5 ng ml⁻¹) (212-12; Peprotech), and recombinant murine granulocyte macrophage colony-stimulating factor (GM-CSF) for BMDC generation (1000 U ml⁻¹) (31503; Peprotech).

Metabolic assays. Metabolic flux assays were performed as described in ref. 3. T cells were activated by 1 μ g ml⁻¹ plate-bound anti-CD3 and anti-CD28 plus ICAM for 36 h, and first division cells were sorted on c-Myc expression. We then plated 5×10^5 cells per replicate with different labelled compounds overnight. Glycolytic flux was determined by measuring the detritiation of [3-³H]glucose, and fatty acid β -oxidation flux was determined by measuring the detritiation of [9,10-³H]palmitic acid. Glutamine uptake was determined by measure the detritiation of [³H]glutamine in lysed cells for 3 h.

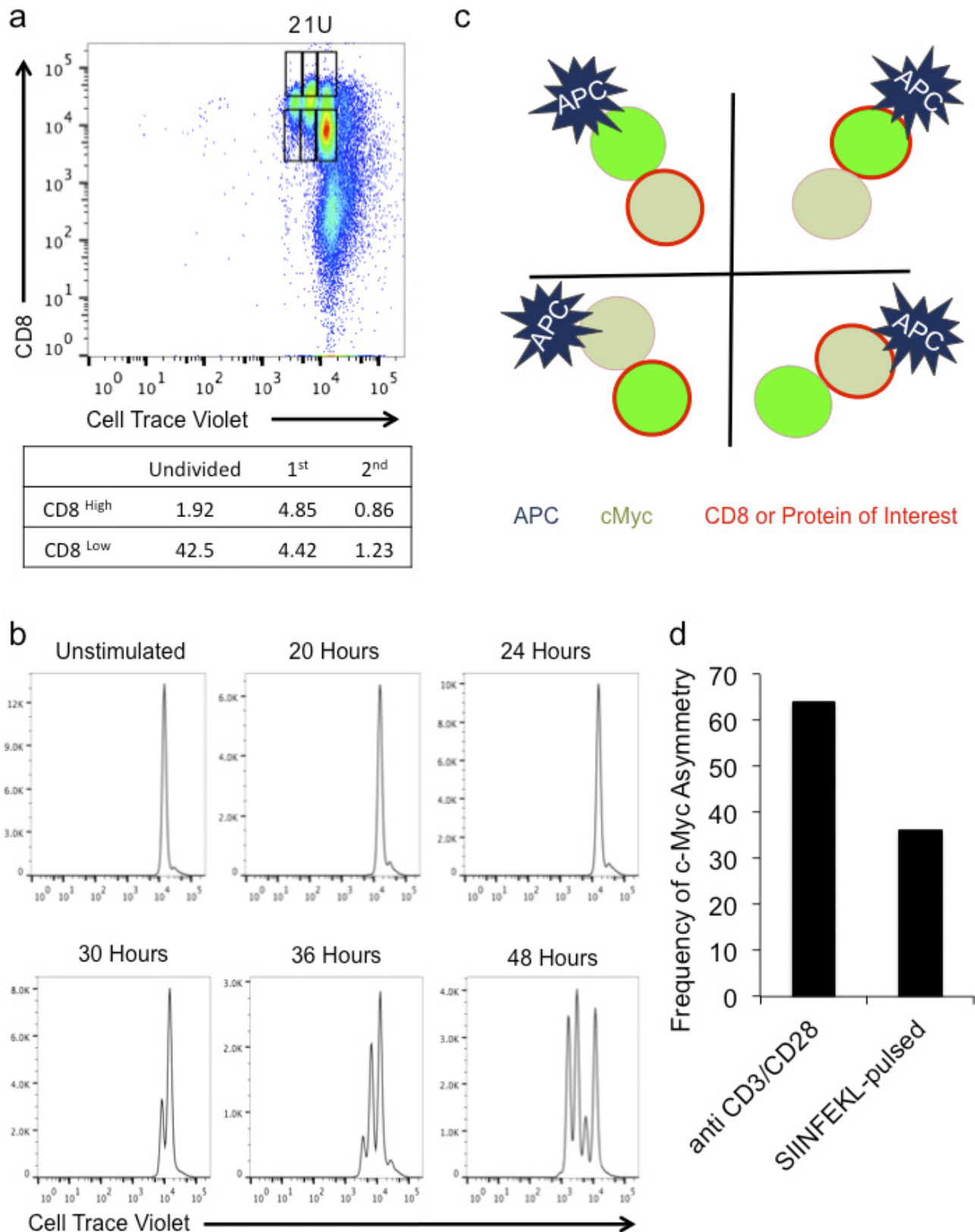
Respiration was measured in intact T cells using the Seahorse XF24 analyser. Primary mouse T lymphocytes were either maintained in resting stage by 5 ng ml⁻¹ IL-7 or activated by 1 μ g ml⁻¹ plate-bound anti-CD3 and anti-CD28 plus ICAM. After 36 h, 1×10^6 resting or active T cells were seeded in plates coated with Cell-Tak (354240; Corning). After 1 h, the plate was loaded into the instrument to determine the oxygen consumption rate and extracellular acidification rate. For glycolytic stress tests, T cells sorted on c-Myc GFP were plated in glucose-free complete DMEM medium. During the course of the assay, cultures were injected with 10 mM glucose (67021; Sigma-Aldrich), 1 μ M oligomycin (04876; Sigma-Aldrich), and 20 mM 2DG.

Unbiased metabolomic profiling was performed by Metabolon, Inc. Primary mouse T lymphocytes were either maintained in a resting state by 5 ng ml⁻¹ IL-7 or activated by 1 μ g ml⁻¹ plate-bound anti-CD3 and anti-CD28 plus ICAM. At the end of culture, activated T cells in the first division were sorted on the basis of c-Myc expression, and 1.6 million T cells were spun down, washed once with cold PBS, and snap-frozen in liquid nitrogen. All the samples were extracted and analysed through ultra-high pressure liquid chromatography mass spectroscopy (UHPLC/MS/MS), as described in Supplementary Methods and Supplementary references 2–4.

Infections. For analysis of c-Myc asymmetry *ex vivo*, 5×10^6 naive CFSE-labelled OT-I T cells were transferred by intravenous (i.v.) injection into the tail vein of congenic wild-type recipients. Twenty-four hours after transfer, recipients were infected i.v. with 5×10^3 colony-forming units of *Listeria monocytogenes* expressing SIINFEKL peptide. On day 3 after infection, spleens of recipient animals were collected and immediately flash-frozen before cryosectioning. Unfixed sections were blocked in TBS containing 2% BSA and 0.05% Tween-20 before incubation with anti-c-Myc and AF405-labelled phalloidin overnight at 4°C. c-Myc was finally detected with Cy3-labelled secondary antibody, and sections were imaged using laser scanning confocal microscopy.

For analysis of the functional consequences of c-Myc asymmetry *in vivo*, CellTrace-Violet-labelled, naive c-Myc–GFP, OT-I cells were activated on plate-bound anti-CD3/CD28/ICAM or SIINFEKL-pulsed BMDCs for 36 h (as described above). CellTrace Violet dilution was used to identify T cells in the first division, which were then sorted on the basis of c-Myc–GFP fluorescence into high and low populations. 2.5×10^5 , 5×10^5 , or 7.5×10^5 CD45.1/CD45.2 c-Myc–GFP high or low cells were transferred i.v. into CD45.1 recipients. Recipient animals were either kept as naive, or infected with 10⁶ EID₅₀ HKx31 influenza expressing SIINFEKL peptide (P. Thomas, St. Jude Children's Research Hospital). Two weeks after transfer, spleens and lymph nodes from half of the animals were examined for the presence of donor cells, as indicated by positive staining for the CD45.2 congenic marker. All remaining recipients were rested for an additional two weeks and then challenged with PR8 influenza expressing SIINFEKL peptide (P. Thomas, St. Jude Children's Research Hospital). Nine days after this challenge infection, spleens and lymph nodes were collected and analysed for the presence of donor cells. For *in vivo* treatment with rapamycin, recipient mice received daily intraperitoneal (i.p.) injections of rapamycin at 75 μ g kg⁻¹ (ref. 30) from days 0–10 after infection. **Statistics.** Indicated statistical analyses were performed using Graphpad Prism software. Four different tests were employed: chi-square goodness of fit was used for imaging experiments with BMDCs, assuming 25% of data points expected in each graph quadrant (representative of four possible staining patterns for the two different proteins). Similarly, for imaging experiments with antibody-activated T cells, a two-tailed binomial test was employed with an expected even distribution (50%) of data points into two quadrants (representative of two possible staining patterns). Linear regression was used on the imaging data to determine whether there was a correlation between the two proteins analysed.

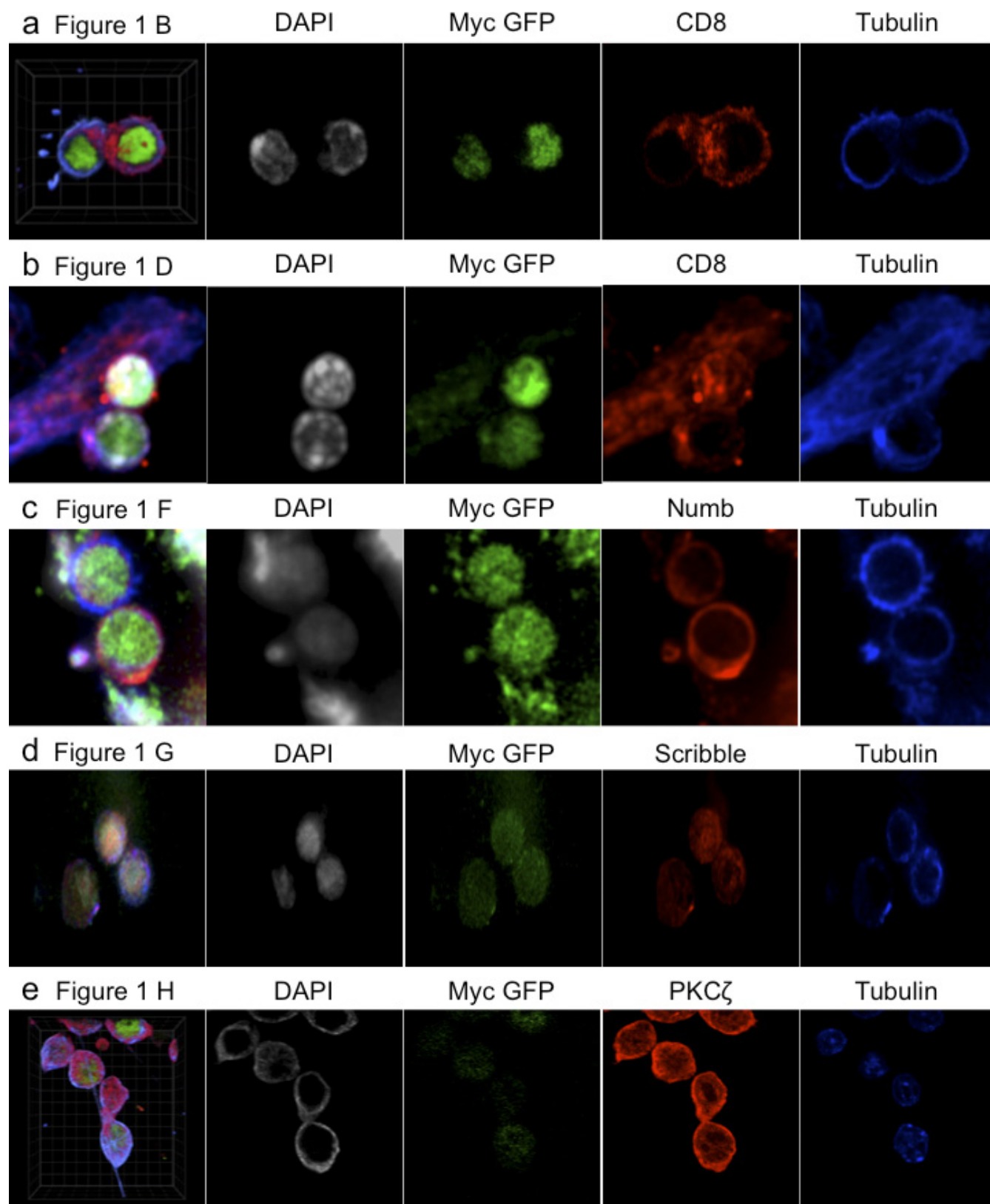
Unpaired Student's *t*-test was used to test whether two means were significantly different from one another when the data are not linked. Paired Student's *t*-test was used to analyse the fluorescence intensity in each daughter of a conjoined pair as in Fig. 1j.



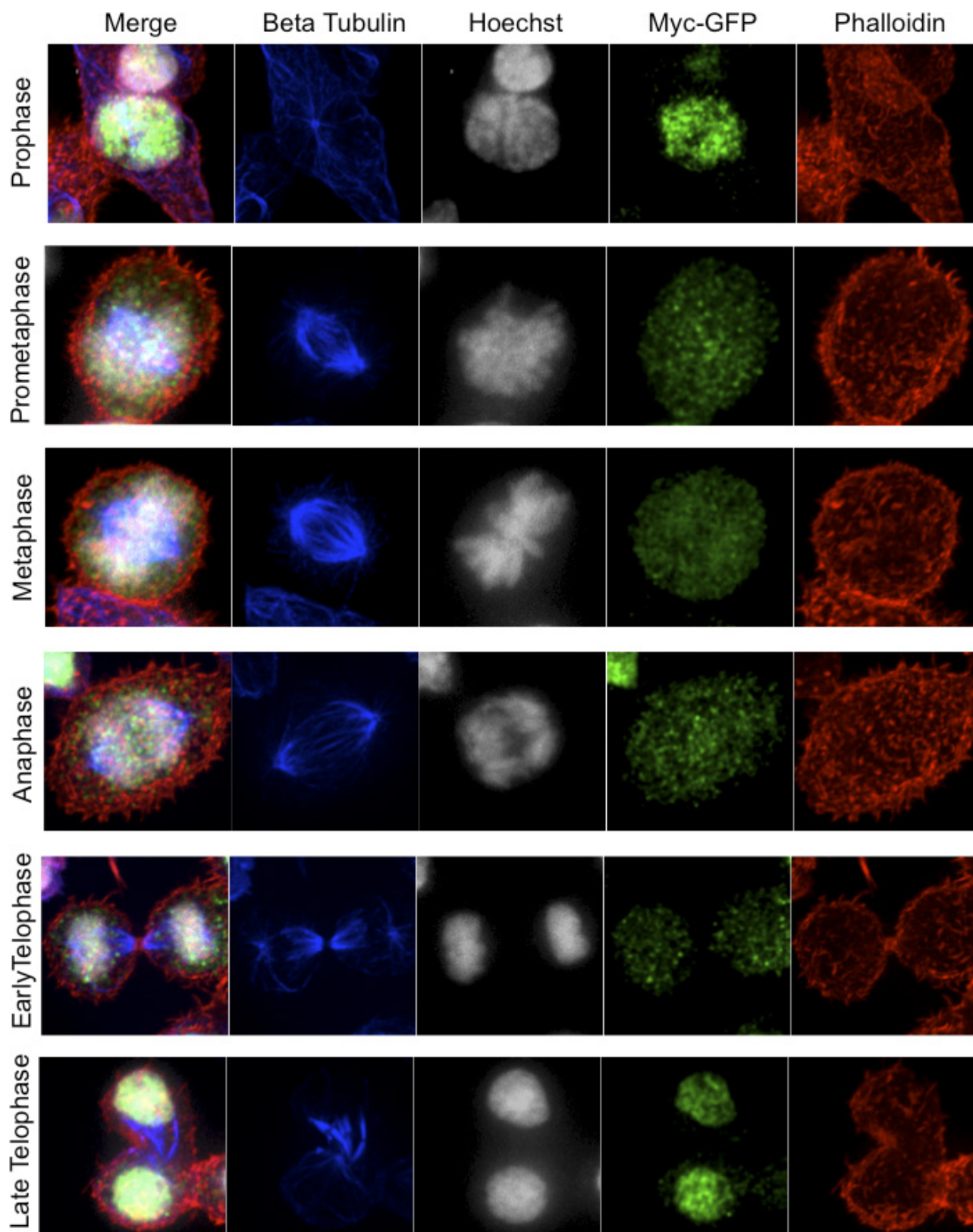
Extended Data Figure 1 | Characterization of c-Myc asymmetry.

a, Gating strategy for CD8^{high} and CD8^{low} cells. CellTrace-Violet-labelled naive T cells were activated with anti-CD3, anti-CD28, and ICAM for 36 h. CD8 high or low cells were identified as activated, undivided (U), first division (1), or second division (2) on the basis of dilution of CellTrace Violet. Frequencies of cells in each population are presented in table below. **b**, Time course of CellTrace Violet dilution in naive, T cells activated with anti-CD3, anti-CD28, and ICAM for the indicated time points. **c**, Schematic representation of staining patterns for interpretation

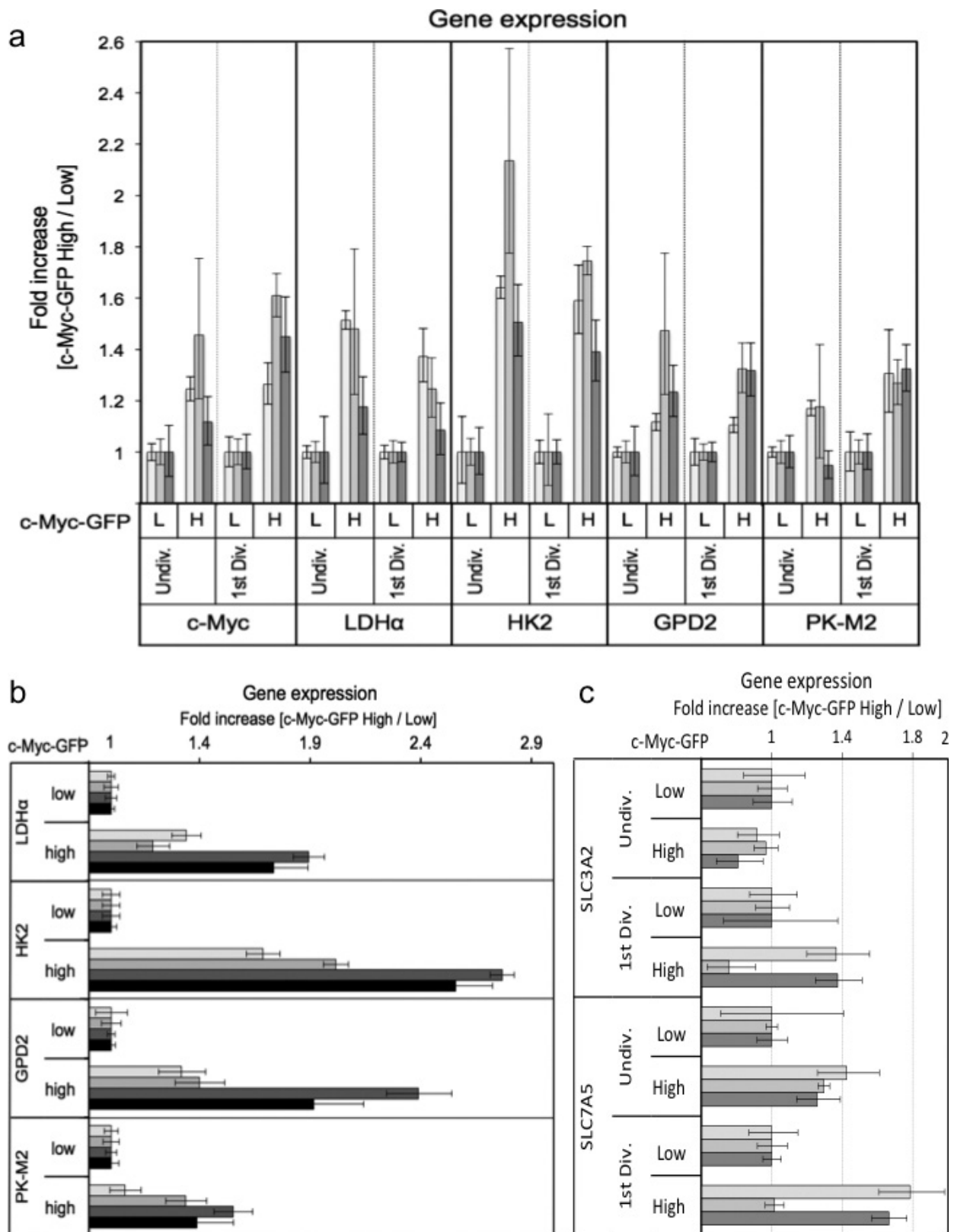
of data represented in Fig. 1c–h. Black lines indicate axes on graphs. **d**, Frequencies of asymmetric cell division were determined by analysing 92 conjoined daughter cells from cultures activated with anti-CD3, anti-CD28, and ICAM and 123 conjoined daughter cells from cultures activated with SIINFEKL-pulsed BMDCs. T cells were analysed for c-Myc-GFP intensity in both daughter cells. Using 1.5-times brightness in one daughter versus the other as a cut off, each pair was assigned either asymmetric or not, and the frequency of asymmetric pairs is plotted.



Extended Data Figure 2 | Representative confocal images and single stains. For data in Fig. 1. a, Fig. 1b, c; b, Fig. 1d, e; c, Fig. 1f; d, Fig. 1g; e, Fig. 1h.

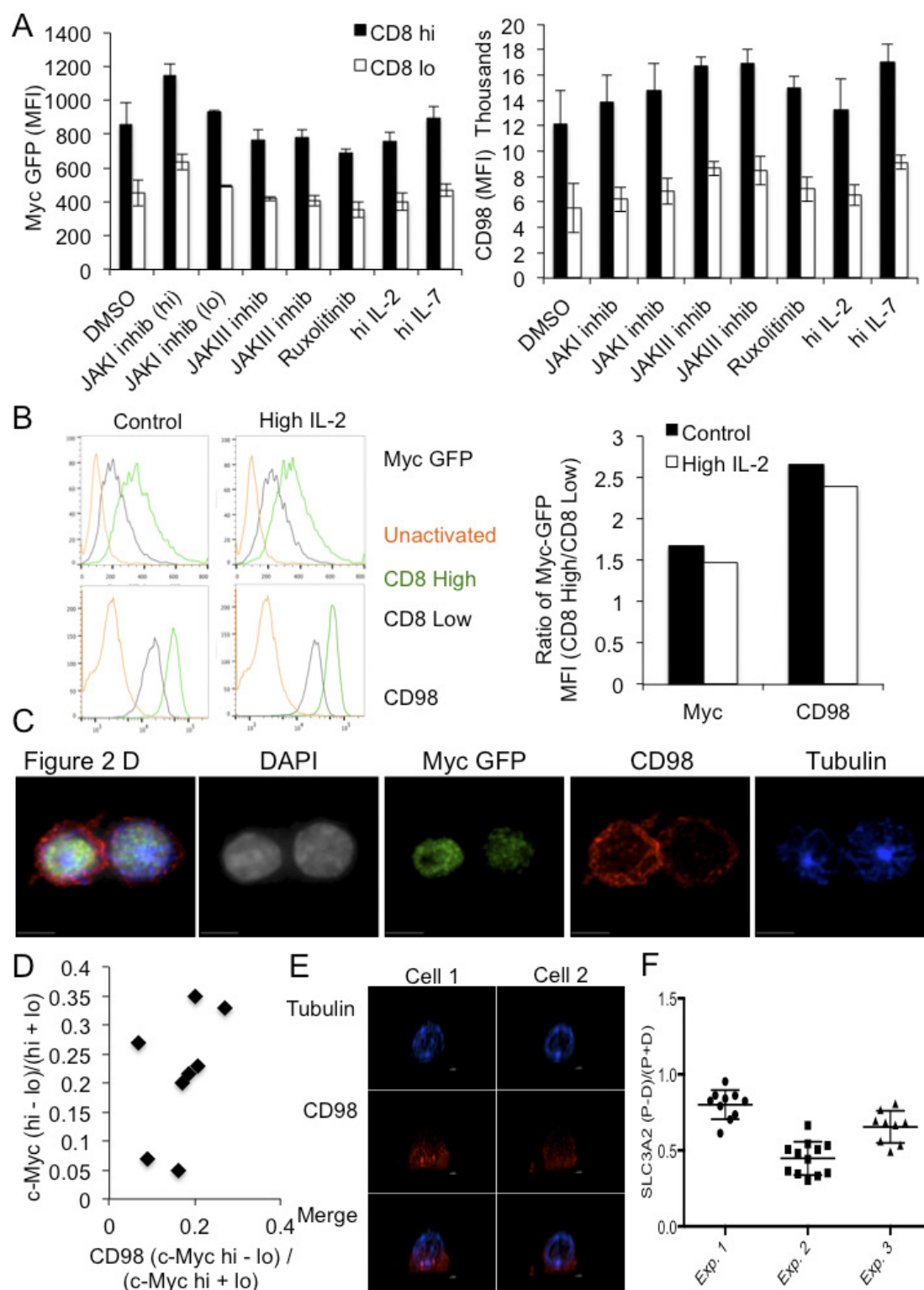


Extended Data Figure 3 | c-Myc expression becomes asymmetric in late telophase. Overlay, β -tubulin (blue), Hoechst 33258 (grey), c-Myc-GFP (green), and phalloidin (red) for each mitotic phase indicated on the left as identified by chromatin and tubulin staining patterns.



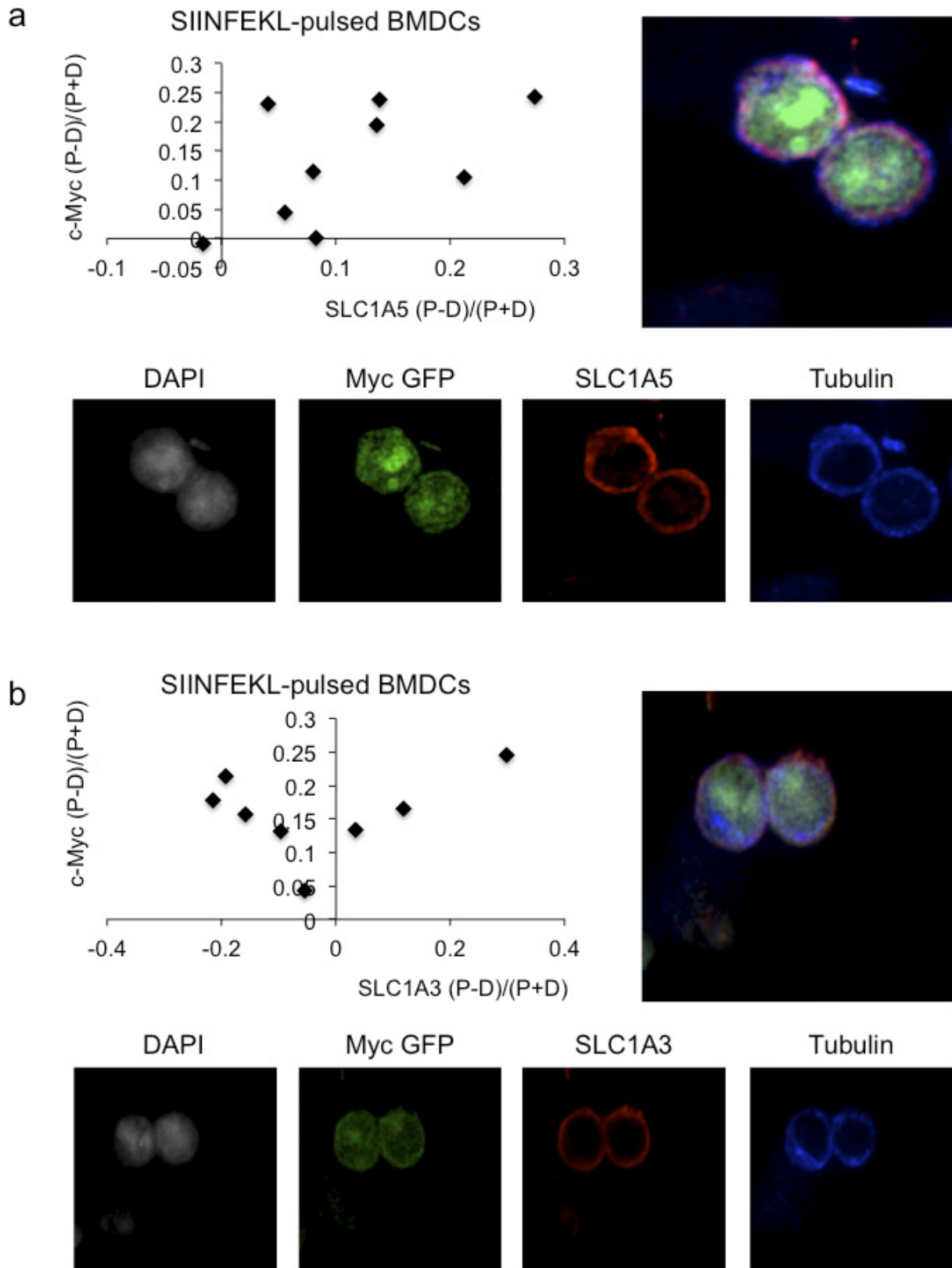
Extended Data Figure 4 | Asymmetric assortment of mRNA for c-Myc target genes in activated T cells. CellTrace-Violet-labelled T cells were stimulated for 36 h with anti-CD3, anti-CD28, and ICAM. **a–c**, First division (**a**, **c**) or undivided (**a–c**) c-Myc^{high} (H) and c-Myc^{low} (L) T cells were sorted on the basis of c-Myc-GFP expression (**b** is undivided cells

only), and RNA was extracted from each population. Fold change in gene expression for the indicated primers is quantified using the $2^{-\Delta\Delta C_t}$ method relative to β -2-microglobulin across three (**a**, **c**) or four (**b**) independent experiments (mean \pm s.d.).



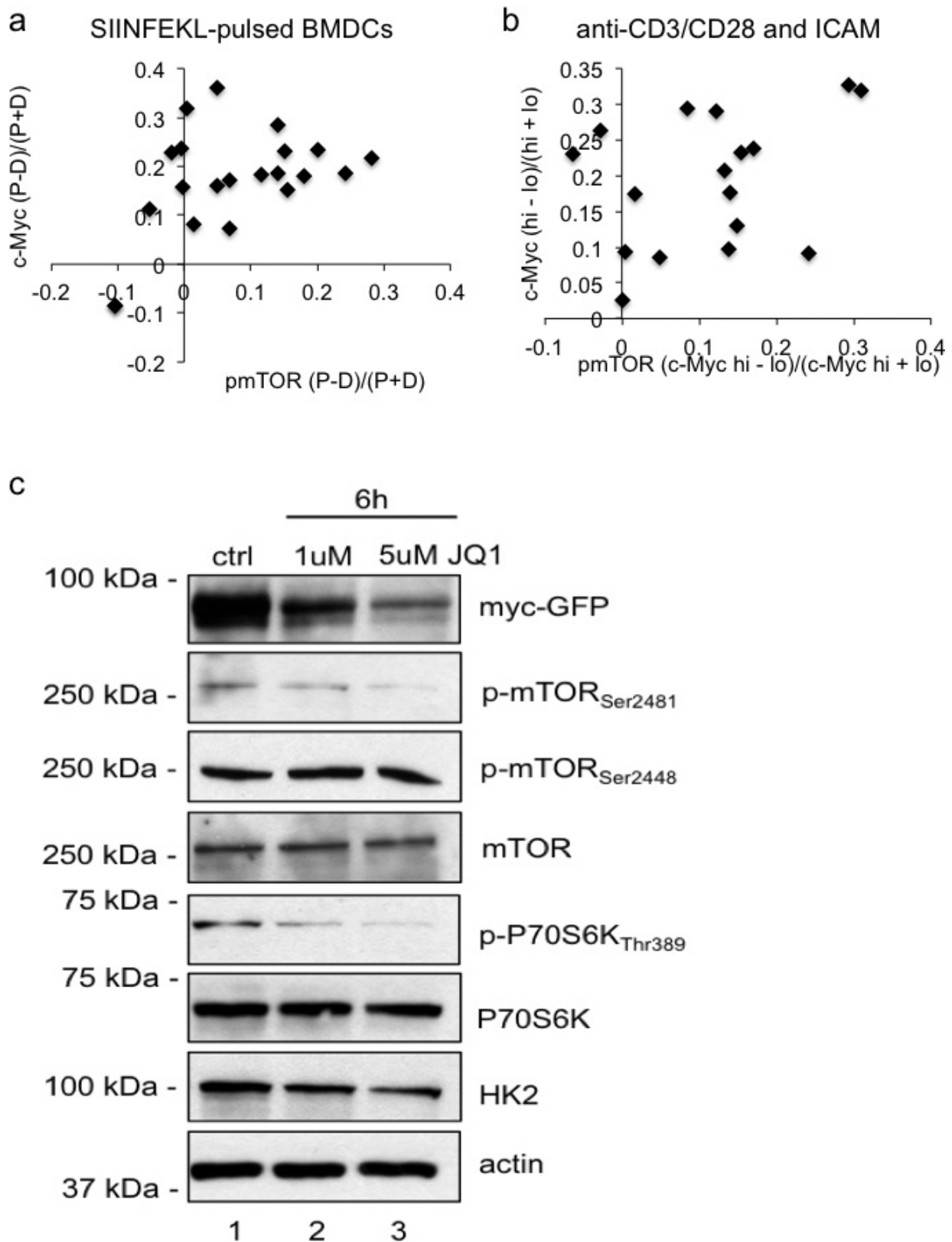
Extended Data Figure 5 | Cytokine signalling does not influence c-Myc asymmetry and asymmetric assortment of CD98. **a**, Mean fluorescent intensity (MFI) of c-Myc-GFP (left panel) or CD98 (right panel) in CD8^{high} (shaded bars) and CD8^{low} (open bars) in the first division after 35 h of activation by anti-CD3, anti-CD8, and ICAM and the indicated treatment for 1 h. All differences were determined significant ($P < 0.05$) by unpaired Student's *t*-test. **b**, Flow cytometric analysis of c-Myc GFP and CD98 in CD8^{high} (green histograms), CD8^{low} (grey histograms), or IL-7 (5 ng ml⁻¹) rested unactivated (gold histograms) T cells in control conditions or activated in the presence of 10 ng ml⁻¹ IL-2 (0–36 h after activation). Representative flow plots are on the left, and quantification of the mean fluorescence intensities of c-Myc and CD98-APC are on the

right. Experiment representative of three independent experiments. **c**, Representative confocal image and single stain images for Fig. 2d, e. OT-I T cells stimulated on SIINFEKL-pulsed BMDCs **d**, Quantification of CD98 and c-Myc asymmetry in T cells stimulated with anti-CD3, anti-CD28, and ICAM for 36 h. 100% concordance of markers ($P = 0.0039$ Two-Tailed Binomial Test); $R^2 = 0.2304$, $P = 0.2304$ linear regression. **e**, Polarization of SLC3A2 in two representative confocal images in antibody-stimulated OT-I Tg T cells (see methods). **f**, Asymmetric index (difference in RFP intensity in proximal (P) and distal (D) sides of cell/sum of RFP intensities in proximal and distal sides of cell) for SLC3A2 staining in activated, undivided CD8 T cells (each point represents an activated T cell). Error bars, mean \pm s.d.



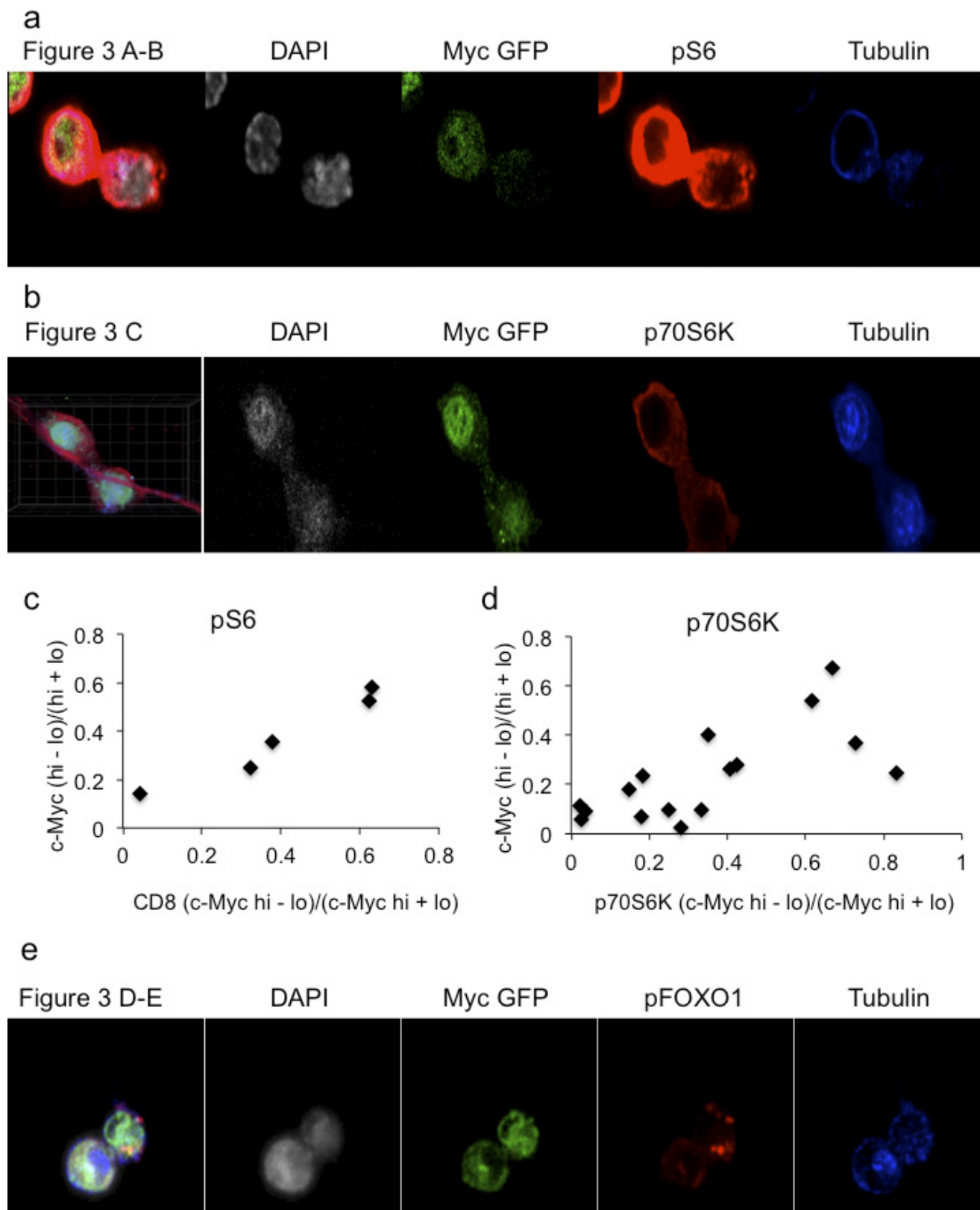
Extended Data Figure 6 | Amino acid transporter SLC1A5, but not SLC1A3, asymmetrically assorts in activated T cells. a, b, Representative confocal image, quantification, and single-stain images for SLC1A5; 88.9% both bright in proximal daughter ($\chi^2 = 19.89$, DF = 3, $P = 0.0002$, chi-square goodness of fit test); $R^2 = 0.2961$, $P = 0.1299$, linear regression (a);

and SLC1A3 100% c-Myc bright in proximal daughter, 62.5% SLC1A3 bright in distal daughter ($\chi^2 = 9$, DF = 3, $P = 0.0293$ chi-square goodness of fit test); $R^2 = 0.07944$, $P = 0.4989$ linear regression (b) for OT-I CD8 T cells co-cultured with SIINFEKL-pulsed BMDCs for 36 h.



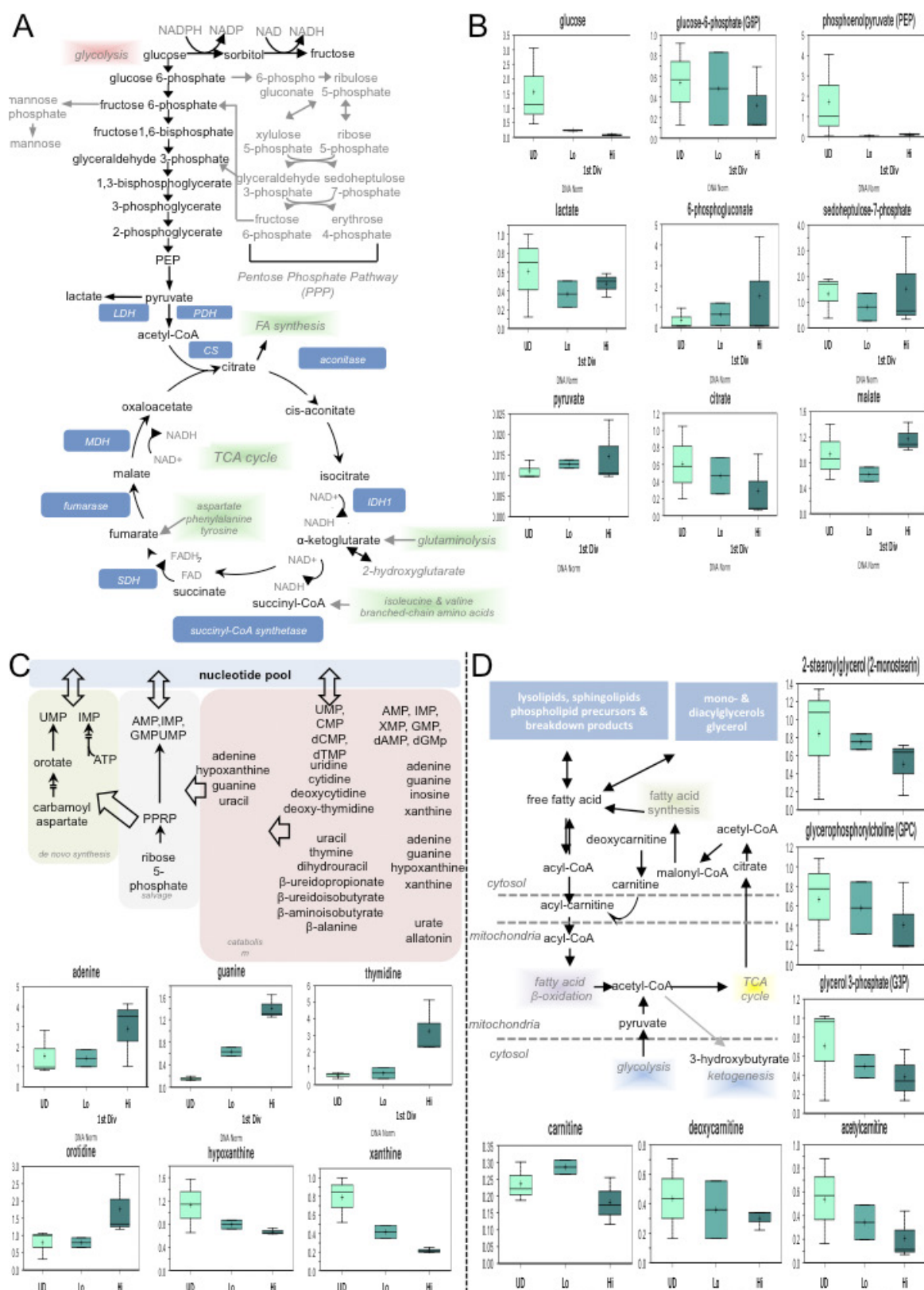
Extended Data Figure 7 | Regulation of p-mTOR and mTORC1 signalling by c-Myc. **a, b**, Quantification of p-mTOR staining and c-Myc-GFP for OT-I CD8 T cells co-cultured with SIINFEKL-pulsed BMDCs 80.9% both bright in proximal daughter ($\chi^2 = 25.67$, $DF = 3$, $P < 0.0001$ chi-square goodness of fit test); $R^2 = 0.2307$, $P = 0.0275$, linear regression (**a**); or T cells stimulated for 36 h with anti-CD3, anti-CD28, and ICAM 82.4% concordance of markers ($P = 0.0013$, two-tailed

binomial test); $R^2 = 0.1204$, $P = 0.1725$, linear regression (**b**). Asymmetry as assessed by fluorescence intensity is expressed as (proximal – distal)/total (**a**), or values from (c-Myc^{high} – c-Myc^{low})/total (**b**). **c**, Western blot analysis of CD8 T cells activated with anti-CD3, anti-CD28, and ICAM for 6 h without treatment (ctrl) or with 1 μ M or 5 μ M of the bromodomain inhibitor JQ1. Data are representative of three independent experiments.



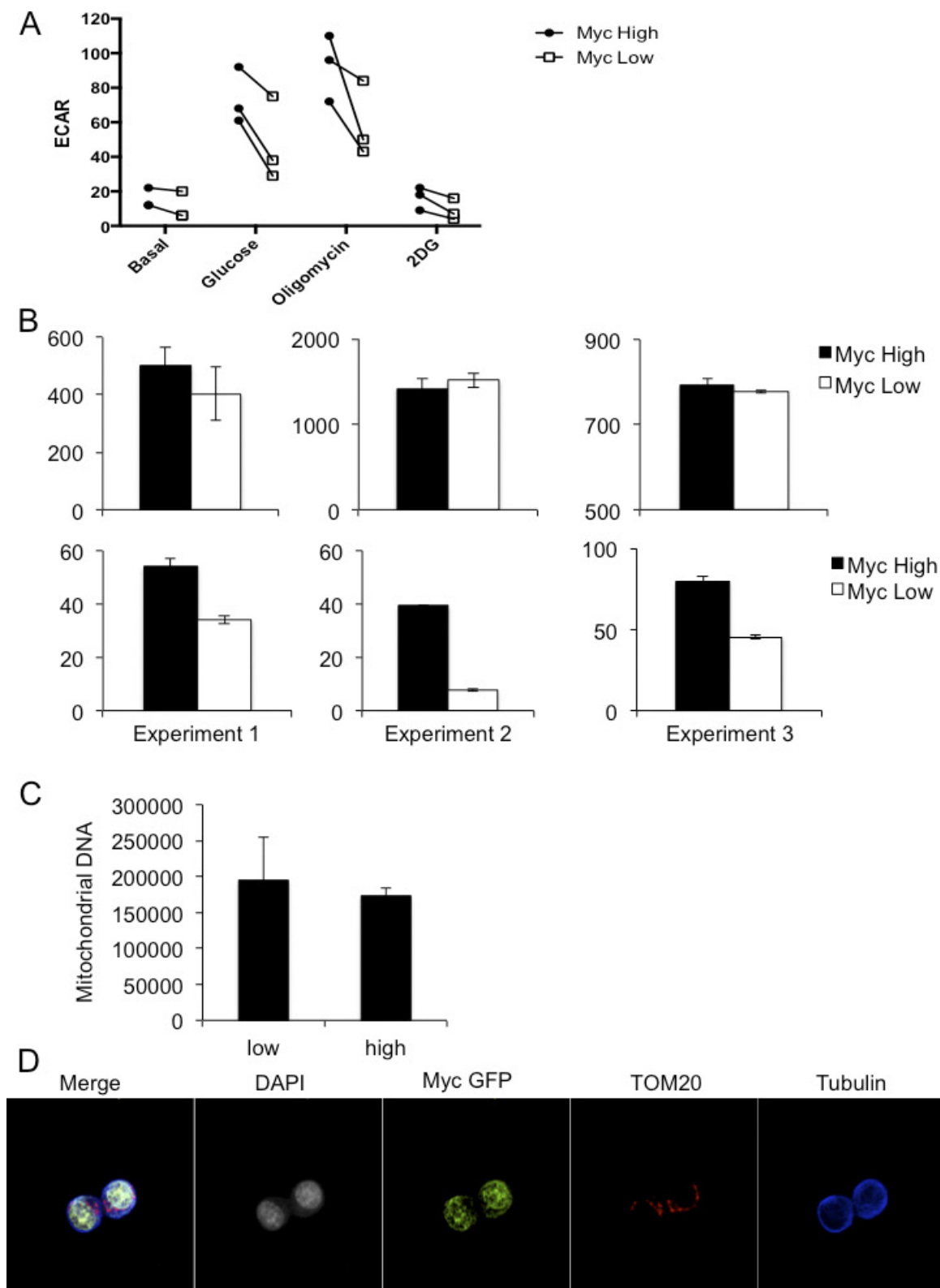
Extended Data Figure 8 | Asymmetric assortment of mTORC1 activity with c-Myc-GFP. a, b, Representative confocal images for quantifications of c-Myc-GFP and pS6 staining corresponding to Fig. 3a, b (a) or c-Myc-GFP and p70S6K corresponding to Fig. 3c (b) in OT-I T cells co-cultured with SIINFEKL-pulsed BMDCs. **c, d,** Quantifications for c-Myc-GFP and pS6 100% both bright in proximal daughter ($\chi^2 = 19.89$, $DF = 3$, $P = 0.0002$, chi-square goodness of fit test); $R^2 = 0.9457$,

$P = 0.0055$, linear regression (c); or p70S6K 100% both bright in proximal daughter ($\chi^2 = 14.14$, $DF = 3$, $P = 0.0027$, chi-square goodness of fit test); $R^2 = 0.4875$, $P = 0.0026$, linear regression (d) staining in T cells stimulated on anti-CD3, anti-CD28, and ICAM for 36 h. **e,** Representative confocal images for quantifications of c-Myc-GFP and pFOXO1 staining corresponding to Fig. 3d, e. T cells stimulated with anti-CD3, anti-CD28, and ICAM for 36 h.



Extended Data Figure 9 | First-division CD8^{high} c-Myc-GFP^{high} CD8 T cells are more glycolytic and exhibit more glutaminolysis and pentose phosphate pathway activity than first division CD8^{low} c-Myc-GFP^{low} CD8 T cells but have decreased FAO. Undivided or first-division antibody-activated T cells were sorted on c-Myc expression. 1.6 million T cells from each group were analysed via UHPLC/MS/MS for unbiased metabolomic profiling performed by Metabolon, Inc. **a–d**,

Metabolic pathway schematic (**a**) was generated by Metabolon Inc., as were graphs (**b**) for select metabolites in pathway, representing data for DNA-normalized data. Metabolic pathway schematics and graphs were also generated by Metabolon Inc. for select metabolites in the pathway, representing data for DNA-normalized data for nucleotide biosynthesis (**c**) and fatty acid oxidation (**d**).



Extended Data Figure 10 | c-Myc^{high} CD8 T cells are more glycolytic than CD8^{low} CD8 T cells without asymmetric distribution of mitochondria. **a**, The extracellular acidification rate (ECAR) for sorted c-Myc^{high} (circles) and c-Myc^{low} (squares) T cells from the first division after 36 h of activation with anti-CD3, anti-CD28, and ICAM as measured by a Seahorse Bioflux analyser during exposure to the indicated compounds are represented. Values are paired across three independent experiments. **b**, The oxygen consumption rate (top panels) and ECAR (bottom panels) for basal respiration in complete RPMI with glucose of sorted c-Myc^{high} (shaded bars) and c-Myc^{low} (open bars) CD8 T cells

from the first-division after 36 h of activation by anti-CD3, anti-CD28, and ICAM, as measured by a Seahorse Bioflux analyser across three independent experiments. **c**, Quantification of mitochondrial DNA in sorted first-division c-Myc low and high OT-I CD8 T cells activated on anti-CD3, anti-CD28 and ICAM for 36 h. Mean + s.d. are plotted for three technical replicates of $n = 2$ mice per group. **d**, Overlay and individual channel images of β -tubulin (blue), DAPI (grey), c-Myc-GFP (green), and TOM20 (red) of an OT-I CD8 T cell activated with anti-CD3, anti-CD28 and ICAM for 36 h.

NOD1 and NOD2 signalling links ER stress with inflammation

A. Marijke Keestra-Gounder^{1*}, Mariana X. Byndloss^{1*}, Núbia Seyffert¹, Briana M. Young¹, Alfredo Chávez-Arroyo¹, April Y. Tsai¹, Stephanie A. Cevallos¹, Maria G. Winter¹, Oanh H. Pham², Connor R. Tiffany¹, Maarten F. de Jong¹, Tobias Kerrinnes¹, Resmi Ravindran², Paul A. Luciw², Stephen J. McSorley², Andreas J. Bäuml^{1*} & Renée M. Tsois^{1*}

Endoplasmic reticulum (ER) stress is a major contributor to inflammatory diseases, such as Crohn disease and type 2 diabetes^{1,2}. ER stress induces the unfolded protein response, which involves activation of three transmembrane receptors, ATF6, PERK and IRE1 α ³. Once activated, IRE1 α recruits TRAF2 to the ER membrane to initiate inflammatory responses via the NF- κ B pathway⁴. Inflammation is commonly triggered when pattern recognition receptors (PRRs), such as Toll-like receptors or nucleotide-binding oligomerization domain (NOD)-like receptors, detect tissue damage or microbial infection. However, it is not clear which PRRs have a major role in inducing inflammation during ER stress. Here we show that NOD1 and NOD2, two members of the NOD-like receptor family of PRRs, are important mediators of ER-stress-induced inflammation in mouse and human cells. The ER stress inducers thapsigargin and dithiothreitol trigger production of the pro-inflammatory cytokine IL-6 in a NOD1/2-dependent fashion. Inflammation and IL-6 production triggered by infection with *Brucella abortus*, which induces ER stress by injecting the type IV secretion system effector protein VceC into host cells⁵, is TRAF2, NOD1/2 and RIP2-dependent and can be reduced by treatment with the ER stress inhibitor tauroursodeoxycholate or an IRE1 α kinase inhibitor. The association of NOD1 and NOD2 with pro-inflammatory responses induced by the IRE1 α /TRAF2 signalling pathway provides a novel link between innate immunity and ER-stress-induced inflammation.

We investigated a potential role for NOD1 and NOD2 in the unfolded protein response (UPR), because this response triggers inflammation by activating IRE1 α , a receptor that recruits TRAF2 to activate NF- κ B (Extended Data Fig. 1a)^{4,6}. NOD1 and NOD2 are two NLRs that contain major TRAF2 binding motifs⁷. Consistent with

previous reports^{8,9}, TRAF2 was required for NF- κ B activation triggered by stimulation with canonical NOD1/2 ligands or mediated by auto-activation of NOD1, NOD2 or the NOD1/2 adaptor protein RIP2 (Extended Data Fig. 1b, c). To investigate whether NOD1 and NOD2 contribute to inflammatory responses during ER stress, we stimulated bone-marrow-derived macrophages (BMDMs) of wild-type (C57BL/6) mice or mice deficient for NOD1 and NOD2 (*Nod1/2*^{-/-} mice) with the ER stress inducer thapsigargin, a specific inhibitor of the SERCA channel¹⁰. Thapsigargin treatment induced the UPR, as indicated by a NOD1/2-independent elevation in transcript levels of *Hspa5* (encoding the ER-resident chaperone GRP78/BiP) and *Chop* (also known as *Ddit3*; Extended Data Fig. 2a, b), two genes controlled by the ER stress sensors ATF6 and PERK (Extended Data Fig. 1a). Thapsigargin treatment also significantly ($P < 0.05$) induced *Il6* transcription (Fig. 1a) and IL-6 synthesis (Fig. 1b), while synthesis of TRAF2 and the chaperones SGT1 and HSP90 remained unchanged (Extended Data Fig. 2c). Both *Il6* transcription and IL-6 synthesis were blunted in BMDMs from *Nod1/2*^{-/-} mice compared to wild-type mice (Fig. 1a, b), which was not due to differences in cell death (Extended Data Fig. 2d). Similarly, IL-6 synthesis induced by the ER stress inducer dithiothreitol was abrogated in BMDMs from *Nod1/2*^{-/-} mice (Extended Data Fig. 2e) which was not due to differences in cell death (Extended Data Fig. 2d). Thapsigargin-induced IL-6 synthesis could be blunted by treatment with the ER stress inhibitor TUDCA (Extended Data Fig. 2f). By contrast, stimulation of BMDMs with the canonical NOD2 ligand muramyl dipeptide (MDP) did not result in increased transcript levels of *Hspa5* and *Chop* (Extended Data Fig. 2a, b), but led to increased synthesis of IL-6 (Extended Data Fig. 2e), which could not be inhibited by TUDCA treatment (Extended Data Fig. 2f). In conclusion, the TUDCA-resistant pathway of IL-6 production induced by stimulation with a canonical

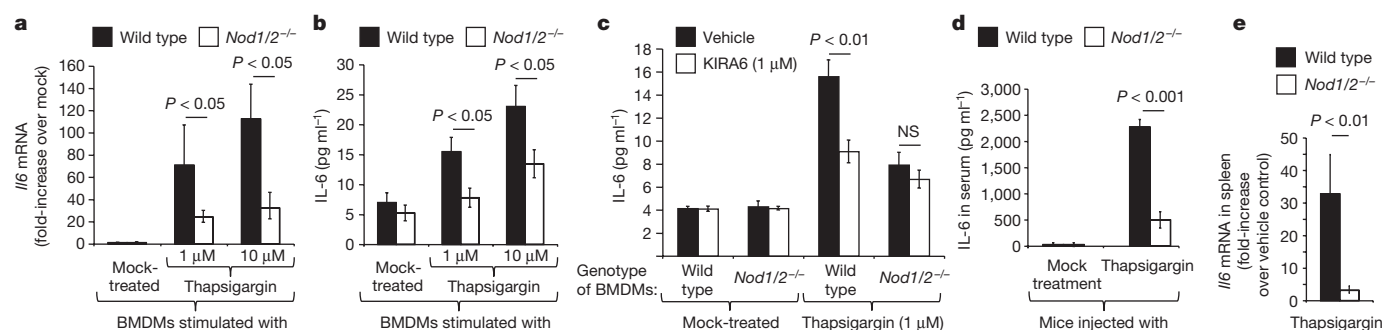


Figure 1 | Thapsigargin induced IL-6 production is dependent on NOD1 and NOD2. **a–c**, BMDMs from *Nod1/2*^{-/-} mice and wild-type littermates ($n = 8$) were stimulated with thapsigargin (**a**, **b**) and/or KIRA6 (**c**) and *Il6* mRNA expression (**a**) and IL-6 protein synthesis (**b**, **c**) were

measured. **d**, **e**, *Nod1/2*^{-/-} mice and wild-type littermates ($n = 7$) were injected with thapsigargin, and IL-6 production in the serum (**d**) and *Il6* mRNA expression in the spleen (**e**) were determined. Data are presented as mean \pm s.e.m.

¹Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, One Shields Ave, Davis, California 95616, USA. ²Center for Comparative Medicine, Schools of Medicine and Veterinary Medicine, University of California at Davis, One Shields Ave, Davis, California 95616, USA.

*These authors contributed equally to this work.

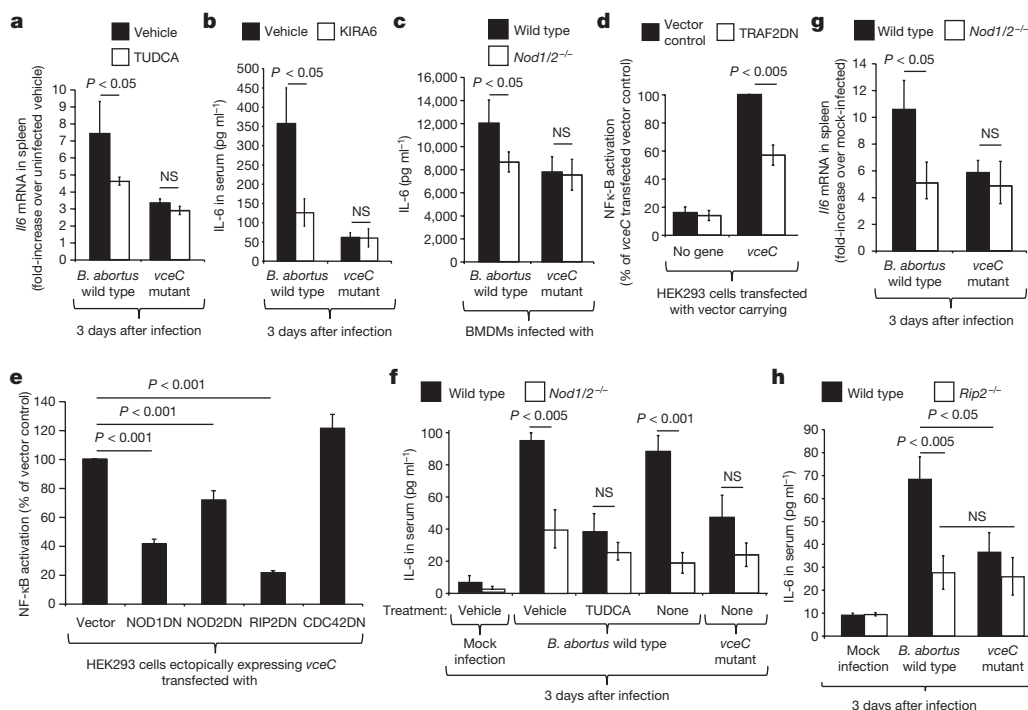


Figure 2 | The ER-stress-associated pro-inflammatory response induced by *B. abortus* VceC is dependent on the NOD1/NOD2 signalling pathway. **a**, *Il6* mRNA expression in the spleen of vehicle-treated or TUDCA-treated mice infected with the indicated *B. abortus* strains (*n* = 5). **b**, **c**, **f**, **g**, **h**, Expression of *Il6* (**g**) or synthesis of IL-6 (**b**, **c**, **f**, **h**)

induced after infection of BMDMs (*n* = 3) (**c**) or mice (*n* = 5; **b**, **f**, **g**, **h**) with the indicated *B. abortus* strains. **d**, **e**, NF-κB activation induced by ectopic expression of VceC in HEK293 cells (*n* = 4) transfected with dominant negative (DN) forms of TRAF2 (**d**), NOD1, NOD2 or RIP2 (**e**). Data are presented as mean ± s.e.m.; NS, not significant.

NOD2 ligand was distinct from the TUDCA-sensitive pathway of IL-6 production induced by thapsigargin (Extended Data Fig. 1a).

The IRE1α kinase domain is important for TRAF2 binding⁴, while the IRE1α RNase domain functions in splicing a 26-nucleotide intronic sequence from the *Xbp1* mRNA¹¹. We thus investigated whether thapsigargin-induced IL-6 synthesis could be blunted by treatment with the IRE1α kinase inhibitor KIRA6 (1-[4-[8-amino-3-tert-butylimidazo[1,5-a]pyrazin-1-yl]naphthalen-1-yl]-3-[3-(trifluoromethyl)phenyl]urea)¹², the IRE1α RNase inhibitor STF-083010 (N-[[2-hydroxynaphthalen-1-yl]methylidene]thiophene-2-sulfonamide)¹³ or the PERK inhibitor GSK2656157¹⁴. Thapsigargin-induced IL-6 synthesis in BMDMs was blunted by treatment with the IRE1α kinase inhibitor KIRA6 (Fig. 1c) which was not due to differences in cell death (Extended Data Fig. 2d). The PERK inhibitor GSK2656157 and the IRE1α RNase inhibitor STF-083010 inhibited expression of *Chop* and *Hspa5*, respectively, but did not reduce thapsigargin-induced IL-6 synthesis (Extended Data Fig. 2g–j). Collectively, these data suggested that the IRE1α kinase branch of the ER stress response was most important for inducing IL-6 synthesis (Extended Data Fig. 1a).

To investigate the *in vivo* relevance of our observations, we injected mice with thapsigargin, which resulted in elevated levels of IL-6, KC (CXCL1) and MIP-1β (CCL4) in the serum (Fig. 1d, Extended Data Fig. 3a–c) and increased transcription of *Il6* in the spleen, liver and kidney (*P* < 0.01) (Fig. 1e, Extended Data Fig. 3d, e) of wild-type mice, while induction of these pro-inflammatory responses was significantly blunted in *Nod1/2^{-/-}* mice. Furthermore, thapsigargin-induced cytokine synthesis could be blunted by treatment with the ER stress inhibitor TUDCA (Extended Data Fig. 3a–c). Collectively, these data suggested that ER stress induced NOD1 and/or NOD2 signalling independently of peptidoglycan and that this pathway was only required for orchestrating the pro-inflammatory branch of the UPR, which leads to IL-6 production (Extended Data Fig. 1a).

ER stress has recently emerged as a response induced during infection with some bacterial pathogens (summarized in ref. 3). *B. abortus*

is an ideal model organism to investigate induction of the UPR *in vivo*, because this pathogen has developed mechanisms to evade recognition by TLRs (reviewed in refs 15, 16). As a result, inflammatory responses induced during brucellosis are almost entirely dependent on a functional type IV secretion system¹⁷, a virulence factor that injects effector proteins, including VceC, into the host cell cytosol¹⁸. VceC translocates to the ER where it binds the ER chaperone BiP and induces an IRE1α-dependent induction of IL-6 production⁵. Profiling of host responses elicited by intraperitoneal infection of mice with *B. abortus* revealed elevated circulating levels of IL-6 (Extended Data Fig. 4a), IL-12p40 (Extended Data Fig. 4b), IFNγ (Extended Data Fig. 4c), KC (Extended Data Fig. 4d), MIP-1β (Extended Data Fig. 4e), G-CSF (Extended Data Fig. 4f) and RANTES (Extended Data Fig. 4g). Blunting of these cytokine levels after treatment with TUDCA confirmed that inflammatory responses were triggered by ER stress (Extended Data Fig. 4). Intraperitoneal infection of mice with *B. abortus* induced elevated transcript levels of *Il6*, but this response was significantly (*P* < 0.05) blunted in mice infected with a *B. abortus vceC* mutant. The VceC-induced pro-inflammatory response was caused by ER stress, because elevated *Il6* transcript levels induced during infection with the *B. abortus* wild type, but not those triggered by the *vceC* mutant, were significantly (*P* < 0.05) blunted after treatment with the ER stress inhibitor TUDCA (Fig. 2a), which was not due to altered bacterial numbers in the spleen (Extended Data Fig. 5a). Furthermore, *Il6* transcript levels induced in the spleen and IL-6 synthesis induced in the serum of mice infected with the *B. abortus* wild type was significantly (*P* < 0.05) blunted by treatment with the IRE1α kinase inhibitor KIRA6 (Extended Data Fig. 5b and Fig. 2b), which was not due to altered bacterial numbers in the spleen (Extended Data Fig. 5c).

We next studied whether VceC-induced ER stress triggered NF-κB activation and IL-6 production through a pathway requiring NOD1 and/or NOD2 activity. *B. abortus* infection elicited significantly (*P* < 0.05) higher production of IL-6 (Fig. 2c) and a significant elevation of *Il6* transcript levels (Extended Data Fig. 5d) in BMDMs from

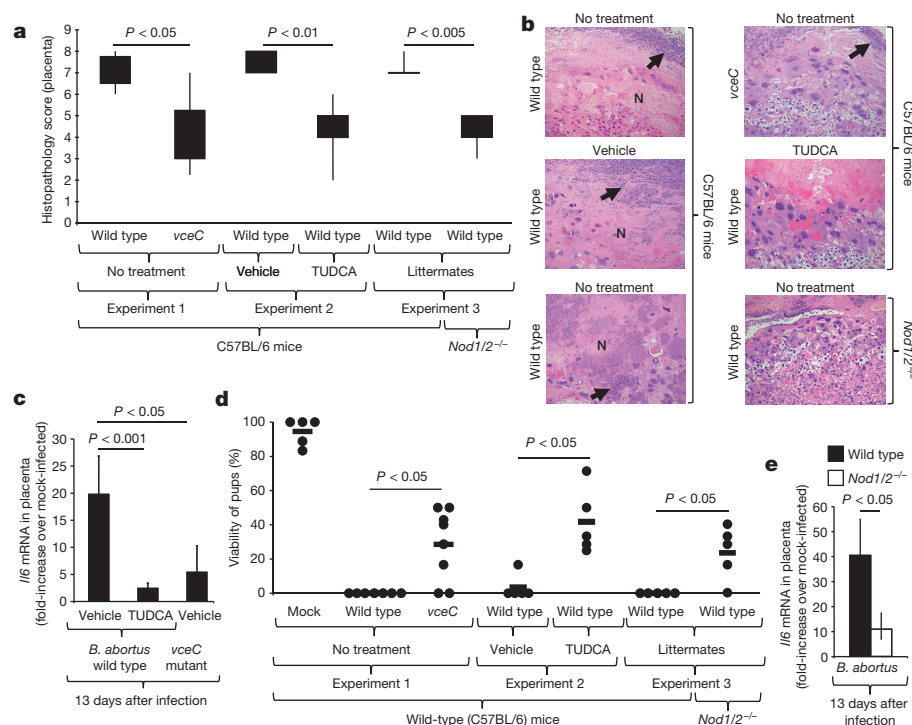


Figure 3 | *B. abortus* VceC-induced ER stress activates the NOD1/NOD2-dependent pro-inflammatory response, thereby contributing to abortion during *B. abortus* infection. **a**, **b**, Histopathology in the placenta of mice infected with *B. abortus*. Whisker plots in **a** represent the second and third quartiles (boxes) and the first and fourth quartiles (lines). **b**, Representative images of histopathological changes. Arrow, neutrophil infiltration;

N, necrosis. **c**, **e**, *Il6* expression in the placenta of vehicle-treated or TUDCA-treated mice infected with the indicated *B. abortus* strains. **d**, Viability of pups. Each dot represents one animal. **e**, *Il6* expression in the placenta of *Nod1/2*^{-/-} mice and wild-type littermates infected with *B. abortus*. The *n* for all experiments is indicated in **d**.

wild-type mice compared to those from *Nod1/2*^{-/-} mice. In contrast, infection with a *vceC* mutant elicited similar responses in BMDMs from wild-type and *Nod1/2*^{-/-} mice. Differences in IL-6 synthesis were not caused by differences in bacterial numbers associated with BMDMs (Extended Data Fig. 5e). Similar results were obtained with BMDMs from *Rip2*^{-/-} mice (Extended Data Fig. 5f). Ectopic expression of VceC in HEK293 cells induced NF- κ B activation, which could be significantly blunted by transfecting cells with dominant negative forms of TRAF2 ($P < 0.005$) (Fig. 2d), NOD1 ($P < 0.001$), NOD2 ($P < 0.001$) or RIP2 ($P < 0.001$) (Fig. 2e). In contrast, transfection with a control protein (a dominant negative form of CDC42) did not reduce VceC-induced NF- κ B activation (Fig. 2e). Collectively, these data suggest that VceC-triggered NF- κ B activation requires NOD1, NOD2 and RIP2 activity.

To investigate the biological significance of NOD1/2-mediated responses to VceC, we infected *Nod1/2*^{-/-} mice and wild-type littermate controls with either the *B. abortus* wild type or a *vceC* mutant. The *B. abortus* wild type elicited significantly higher levels of circulating IL-6 (Fig. 2f) and splenic *Il6* gene transcription (Fig. 2g) in wild-type mice compared to *Nod1/2*^{-/-} mice, although the latter animals did not exhibit reduced bacterial tissue load (Extended Data Fig. 5g). TUDCA treatment blunted IL-6 synthesis in wild-type mice but not in *Nod1/2*^{-/-} mice (Fig. 2f). Responses triggered by the *vceC* mutant were not significantly different in wild-type and *Nod1/2*^{-/-} mice (Fig. 2f, g). Similarly, the *B. abortus* wild type elicited significantly higher IL-6 synthesis (Fig. 2h) and splenic *Il6* expression (Extended Data Fig. 5h) in wild-type mice compared to *Rip2*^{-/-} mice, which was not due to differences in bacterial load (Extended Data Fig. 5i). These data supported the idea that the NOD1/NOD2/RIP2-signalling pathway was required for the VceC-dependent induction of pro-inflammatory responses during *B. abortus* infection *in vivo*.

B. abortus generally causes mild chronic inflammation during its persistence in the tissue of experimentally infected mice or within its

natural bovine reservoir. However, in pregnant cows, the pathogen causes severe acute inflammation in the placenta, which leads to abortion. We next wanted to investigate the contribution of ER stress to acute placentitis using the pregnant mouse model¹⁹. Bacterial numbers in the placenta of mice infected with *B. abortus* wild type at 5 days of pregnancy increased to 10⁹ colony-forming units (CFU) per g tissue by day 13 after infection (corresponding to day 18 of gestation) (Extended Data Fig. 6a), which was accompanied by a significant ($P < 0.05$) increase in *Il6* transcript levels in the placenta (Extended Data Fig. 6b) and acute placentitis (Extended Data Figs 6c–e). This time point was chosen to investigate a potential role for VceC in causing placentitis. Although the *B. abortus* wild type and the *vceC* mutant were recovered in similar numbers from the placenta (Extended Data Fig. 7a), the severity of placentitis (Fig. 3a, b) and the *Il6* transcript levels in the placenta (Fig. 3c) were significantly ($P < 0.05$) reduced in mice infected with the *vceC* mutant. Notably, while none of the pups survived an infection with the *B. abortus* wild type, in mice infected with the *vceC* mutant approximately 30% of the pups were viable at day 18 of gestation (Fig. 3d). These data suggest that VceC-induced inflammation contributes to both placentitis and reduction of pup viability in this model.

Injecting *B. abortus* infected pregnant mice with the ER stress inhibitor TUDCA decreased *Il6* transcript levels in the placenta ($P < 0.001$) (Fig. 3c), reduced the severity of placentitis ($P < 0.01$) (Fig. 3a, b), and increased the viability of pups to approximately 40% (Fig. 3d), while recovery of bacteria from organs remained similar (Extended Data Fig. 7b). To examine whether the NOD1/2 signalling pathway contributed to placentitis, pregnant *Nod1/2*^{-/-} mice and wild-type littermate control mice were infected with *B. abortus*. Compared to infection of pregnant wild-type mice, *B. abortus* infection of pregnant *Nod1/2*^{-/-} mice produced diminished transcript levels of *Il6* in the placenta ($P < 0.05$) (Fig. 3e), reduced the severity of placentitis ($P < 0.005$) (Fig. 3a, b), and increased the viability of pups to approximately 30% (Fig. 3d), while recovery of bacteria from organs remained unchanged

(Extended Data Fig. 7c). To conclude, our data suggest that VceC triggers ER stress, thereby inducing the NOD1/2-dependent pro-inflammatory branch of the UPR, which contributes to abortion during *B. abortus* infection.

To investigate whether the proposed pathway was induced during infections with other pathogens, we infected HeLa cells with *Chlamydia muridarum*, a known ER stress inducer²⁰. Infection of HeLa cells with *C. muridarum* or treatment with thapsigargin induced *Il6* mRNA expression, which could be blunted by treatment with the IRE1 α kinase inhibitor KIRA6 (Extended Data Fig. 8). Consistent with our model (Extended Data Fig. 1a), KIRA6 treatment did not blunt *Il6* expression induced by stimulation of HeLa cells with the NOD2 ligand MDP. Inhibition of NOD1 and NOD2 signalling by transfecting HeLa cells with a dominant negative form of RIP2 blunted *Il6* expression induced by *C. muridarum* infection, thapsigargin treatment or stimulation with MDP (Extended Data Fig. 8). These results suggest that the NOD1/NOD2/RIP2 signalling pathway also detects ER stress generated during *Chlamydia* infection.

NOD1 and NOD2 are traditionally viewed as cytosolic sensors of bacterial peptidoglycan fragments. A clue that NOD2 may perform functions in addition to sensing canonical bacterial peptidoglycan ligands comes from the finding that this PRR activates innate immune responses during infection with influenza virus²¹, a pathogen that triggers ER stress in infected host cells²². Our results suggest that pro-inflammatory responses induced by ER stress are mediated through a NOD1/2-dependent, TUDCA/KIRA6-sensitive pathway, which differs from the TUDCA/KIRA6-resistant pathways induced by bacterial peptidoglycan fragments (Extended Data Fig. 1a). This finding provides a new link between NOD1, NOD2 and inflammatory diseases involving ER stress, such as Crohn disease, ulcerative colitis, obesity and type 2 diabetes^{1,2}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 April 2015; accepted 3 March 2016.

Published online 23 March 2016.

1. Kaser, A. *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008).
2. Montane, J., Cadavez, L. & Novials, A. Stress and the inflammatory process: a major cause of pancreatic cell death in type 2 diabetes. *Diabetes Metab. Syndr. Obes.* **7**, 25–34, (2014).
3. Celli, J. & Tsolis, R. M. Bacteria, the endoplasmic reticulum and the unfolded protein response: friends or foes? *Nature Rev. Microbiol.* (2014).
4. Urano, F. *et al.* Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. *Science* **287**, 664–666 (2000).
5. de Jong, M. F. *et al.* Sensing of bacterial type IV secretion via the unfolded protein response. *MBio* **4**, e00418–12 (2013).
6. Kaneko, M., Niinuma, Y. & Nomura, Y. Activation signal of nuclear factor- κ B in response to endoplasmic reticulum stress is transduced via IRE1 and tumor necrosis factor receptor-associated factor 2. *Biol. Pharm. Bull.* **26**, 931–935 (2003).

7. Schneider, M. *et al.* The innate immune sensor NLRC3 attenuates Toll-like receptor signaling via modification of the signaling adaptor TRAF6 and transcription factor NF- κ B. *Nature Immunol.* **13**, 823–831 (2012).
8. McCarthy, J. V., Ni, J. & Dixit, V. M. RIP2 is a novel NF- κ B-activating and cell death-inducing kinase. *J. Biol. Chem.* **273**, 16968–16975 (1998).
9. Li, L. *et al.* TRIP6 is a RIP2-associated common signaling component of multiple NF- κ B activation pathways. *J. Cell Sci.* **118**, 555–563 (2005).
10. Lytton, J., Westlin, M. & Hanley, M. R. Thapsigargin inhibits the sarcoplasmic or endoplasmic reticulum Ca-ATPase family of calcium pumps. *J. Biol. Chem.* **266**, 17067–17071 (1991).
11. Calton, M. *et al.* IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415**, 92–96 (2002).
12. Ghosh, R. *et al.* Allosteric inhibition of the IRE1 α RNase preserves cell viability and function during endoplasmic reticulum stress. *Cell* **158**, 534–548 (2014).
13. Papandreou, I. *et al.* Identification of an IRE1 α endonuclease specific inhibitor with cytotoxic activity against human multiple myeloma. *Blood* **117**, 1311–1314 (2011).
14. Atkins, C. *et al.* Characterization of a novel PERK kinase inhibitor with antitumor and antiangiogenic activity. *Cancer Res.* **73**, 1993–2002 (2013).
15. Tsolis, R. M., Young, G. M., Solnick, J. V. & Bauml, A. J. From bench to bedside: stealth of enteroinvasive pathogens. *Nature Rev. Microbiol.* **6**, 883–892 (2008).
16. Martirosyan, A., Moreno, E. & Gorvel, J. P. An evolutionary strategy for a stealthy intracellular *Brucella* pathogen. *Immunol. Rev.* **240**, 211–234 (2011).
17. Roux, C. M. *et al.* *Brucella* requires a functional type IV secretion system to elicit innate immune responses in mice. *Cell. Microbiol.* **9**, 1851–1869 (2007).
18. de Jong, M. F., Sun, Y. H., den Hartigh, A. B., van Dijk, J. M. & Tsolis, R. M. Identification of VceA and VceC, two members of the VjbR regulon that are translocated into macrophages by the *Brucella* type IV secretion system. *Mol. Microbiol.* **70**, 1378–1396 (2008).
19. Kim, S. *et al.* Interferon- γ promotes abortion due to *Brucella* infection in pregnant mice. *BMC Microbiol.* **5**, 22 (2005).
20. Derre, I. Chlamydiae interaction with the endoplasmic reticulum: contact, function and consequences. *Cell. Microbiol.* **17**, 959–966 (2015).
21. Sabbah, A. *et al.* Activation of innate immune antiviral responses by Nod2. *Nature Immunol.* **10**, 1073–1080 (2009).
22. Roberson, E. C. *et al.* Influenza induces endoplasmic reticulum stress, caspase-12-dependent apoptosis, and c-Jun N-terminal kinase-mediated transforming growth factor- β release in lung epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **46**, 573–581 (2012).

Acknowledgements Work in R.M.T.'s laboratory is supported by US Public Health Service (USPHS) Grants AI112258 and AI109799. Work in A.J.B.'s laboratory was supported by USPHS Grants AI044170, AI076246 and AI096528. Work in S.J.M.'s laboratory is supported by USPHS Grants AI076278 and AI117303. S.A.C. was supported by USPHS Grant GM056765. A.M.K.-G. is supported by the American Heart Association Grant 12SDG12220022. N.S. was supported by a CAPES Science without Borders fellowship.

Author Contributions A.M.K.-G. and M.X.B. performed and analysed the experiments. R.R., P.A.L., O.H.P., A.Y.T., S.A.C., C.R.T., N.B.S., B.M.Y., A.C.-A., T.K., M.F.d.J. and M.G.W. performed experiments. A.M.K.-G., M.X.B., S.J.M., A.J.B. and R.M.T. were responsible for the overall study design and for writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.T. (rmtsolis@ucdavis.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were blinded to allocation of mice for assessment of histopathology and readouts of inflammation.

Bacterial strains, tissue culture cells and culture conditions. *E. coli* strains were routinely cultured aerobically at 37°C in lysogeny broth (LB) and on LB agar plates. *B. abortus* was cultured in tryptic soy broth or on tryptic soy agar (TSA) plates. *Chlamydia muridarum* strain Nigg II was purchased from ATCC (Manassas, VA). Bacteria were cultured in HeLa 229 cells in DMEM supplemented with 10% FBS. Elementary bodies (EBs) were purified by discontinuous density gradient centrifugations as described previously²³ and stored at -80°C. The HEK293 cell line was maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% FBS at 37°C in a 5% CO₂ atmosphere.

Transfections. HEK293 cells (ATCC CRL-1573) were obtained from ATCC and were grown in a 48-well tissue culture plates in DMEM containing 10% FBS until ~40% of confluency was reached. HEK293 cells were transfected with a total of 250 ng of plasmid DNA per well, consisting of 25 ng of the reporter construct pNF-κB-luc, 25 ng of the normalization vector pTK-LacZ, and 200 ng of the different combinations of mammalian expression vectors carrying the indicated gene (empty control vector, pCMV-HA-VceC⁵, pCMV-HA-TRAF2DN (this study), hNOD1-3×Flag, hNOD2-3×Flag, pCMV-HA-hRip2, hNOD1DN-3×Flag, hNOD2DN-3×Flag or pCMV-HA-Rip2DN²⁴ and pCMV-myc-CDC42DN²⁵). The dominant-negative form of TRAF2, lacking an amino-terminal RING finger domain²⁶, was PCR amplified from cDNA prepared from HEK293 cells and cloned into the mammalian expression vector pCMV-HA (BD Biosciences Clontech). Forty-eight hours after transfection, cells were lysed either without any treatment, or stimulated with C12-iE-DAP (1,000 ng ml⁻¹, InvivoGen) and MDP (10 μg ml⁻¹, InvivoGen). After five hours of treatment the cells were lysed and analysed for β-galactosidase and luciferase activity (Promega). FuGene HD (Roche) was used as a transfection reagent according to the manufacturer's instructions. Cell lines were monitored for mycoplasma contamination.

Bone-marrow-derived macrophage infection. Bone-marrow-derived macrophages (BMDMs) were differentiated from bone marrow precursors from femur and tibiae of C57BL/6 mice obtained from The Jackson Laboratory (Bar Harbor, ME), *Nod1*^{+/-}*Nod2*^{+/-} (wild-type littermates) and *Nod1*^{-/-}*Nod2*^{-/-} (NOD1/NOD2-deficient) mice (generated at UC Davis) as described previously²⁷. For BMDM experiments, 24-well microtitre plates were seeded with macrophages at a concentration of 5 × 10⁵ cells per well in 0.5 ml of RPMI media (Invitrogen, Grand Island, NY) supplemented with 10% FBS and 10 mM L-glutamine (complete RPMI) and incubated for 48 h at 37°C in 5% CO₂. BMDMs were stimulated with C12-iE-DAP (1,000 ng ml⁻¹, InvivoGen), MDP (10 μg ml⁻¹, InvivoGen), thapsigargin (1 μM and 10 μM, Sigma-Aldrich), dithiothreitol (DTT) (1 mM, Sigma-Aldrich), and LPS (10 ng ml⁻¹, InvivoGen) with or without pre-treatment (30 min) of the cells with IRE1α kinase inhibitor KIRA6 (1 μM, Calbiochem), IRE1α endonuclease inhibitor STF-083010 (50 μM, Sigma-Aldrich), PERK inhibitor GSK2656157 (500 nM, Calbiochem) and tauroursodeoxycholate TUDCA (200 μM, Sigma-Aldrich) in the presence of 1 ng ml⁻¹ of recombinant mouse IFNγ (BD Bioscience, San Jose, CA). After 24 h of stimulation, samples for ELISA and gene expression analysis were collected as described below. Preparation of the *B. abortus* wild-type strain 2308 and the *ΔvceC* mutant inoculum and BMDM infection was performed as previously described²⁷. Approximately 5 × 10⁷ bacteria in 0.5 ml of complete RPMI were added to each well containing 5 × 10⁵ BMDMs. Microtitre plates were centrifuged at 210g for 5 min at room temperature in order to synchronize infection. Cells were incubated for 20 min at 37°C in 5% CO₂, and free bacteria were removed by three washes with PBS, and the zero-time-point sample was taken as described below. After the PBS wash, complete RPMI plus 50 mg ml⁻¹ gentamicin and 1 ng ml⁻¹ of recombinant mouse IFNγ (BD Bioscience, San Jose, CA) was added to the cells, and incubated at 37°C in 5% CO₂. For cytokine production assays, supernatant for each well was sampled at 24 h after infection. In order to determine bacterial survival, the medium was aspirated at the time point described above, and the BMDMs were lysed with 0.5 ml of 0.5% Tween 20, followed by rinsing each well with 0.5 ml of PBS. Viable bacteria were quantified by serial dilution in sterile PBS and plating on TSA. For gene expression assays, BMDMs were suspended in 0.5 ml of TRI-reagent (Molecular Research Center, Cincinnati) at the time points described above and kept at -80°C until further use. At least three independent assays were performed with triplicate samples, and the standard error of the mean for each time point was calculated.

Animal experiments. All mouse experiments were approved by the Institutional Animal Care and Use Committees at the University of California, Davis, and were conducted in accordance with institutional guidelines. Sample sizes were determined based on experience with infection models and were calculated to use the minimum number of animals possible to generate reproducible results. C57BL/6

wild-type mice and *Rip2*^{-/-} mice (The Jackson Laboratory), *Nod1*^{+/-}*Nod2*^{+/-} (wild-type littermates) and *Nod1*^{-/-}*Nod2*^{-/-} (NOD1/NOD2-deficient) mice (generated at UC Davis) were injected intraperitoneally (i.p.) with 100 μl of 2.5 mg per kg body weight of thapsigargin (Sigma-Aldrich) at 0 and 24 h, and 4 h after the second injection the mice were euthanized and serum and tissues collected for gene expression analysis and detection of cytokines. Where indicated, mice were treated i.p. at 12 h before the first thapsigargin dose and 12 h before the second thapsigargin dose with the ER stress inhibitor TUDCA (250 mg per kg body weight).

Female and male C57BL/6, *Nod1*^{+/-}*Nod2*^{+/-}, *Nod1*^{-/-}*Nod2*^{-/-} mice, and *Rip2*^{-/-} mice aged 6–8 weeks, were held in micro-isolator cages with sterile bedding and irradiated feed in a biosafety level 3 laboratory. Groups of five mice were inoculated i.p. with 0.2 ml of PBS containing 5 × 10⁵ CFU of *B. abortus* 2308 or its isogenic mutant *ΔvceC*, as previously described²⁸. At 3 days post-infection, mice were euthanized by CO₂ asphyxiation and their serum and spleens were collected aseptically at necropsy. The spleens were homogenized in 2 ml of PBS, and serial dilutions of the homogenate were plated on TSA for enumeration of CFU. Spleen samples were also collected for gene expression analysis as described below. When necessary, mice were treated i.p. at day one and two post-infection with a daily dose of 250 mg per kg body weight of the ER stress inhibitor TUDCA (Sigma-Aldrich), or 10 mg per kg body weight of the IRE1α kinase inhibitor KIRA6 (Calbiochem) or vehicle control.

For the placatitis mouse model, C57BL/6, *Nod1*^{+/-}*Nod2*^{+/-} and *Nod1*^{-/-}*Nod2*^{-/-} mice, aged 8–10 weeks, were held in micro-isolator cages with sterile bedding and irradiated feed in a biosafety level 3 laboratory. Female *Nod1*^{+/-}*Nod2*^{+/-} mice were mated with male C57BL/6 mice (control mice) and female *Nod1*^{-/-}*Nod2*^{-/-} mice were mated with male *Nod1*^{-/-}*Nod2*^{-/-} mice (NOD1/NOD2-deficient), and pregnancy was confirmed by presence of a vaginal plug. At 5 days of gestation, groups of pregnant mice were mock infected or infected i.p. with 1 × 10⁵ CFU of *Brucella abortus* 2308 or its isogenic mutant *ΔvceC* (day 0). At 3, 7 and 13 days after infection mice were euthanized by CO₂ asphyxiation and the spleen and placenta of dams were collected aseptically at necropsy. At day 13 after infection (corresponding to day 18 of gestation), viability of pups was evaluated based on the presence of fetal movement and heartbeat, and fetal size and skin colour. Fetuses were scored as viable if they exhibited movement and a heartbeat, visible blood vessels, bright pink skin, and were of normal size for their gestational period. Fetuses were scored as non-viable if fetal movement, heartbeat, and visible blood vessels were absent, skin was pale or opaque, and their size for gestational period or compared to littermates was small, or they showed evidence of fetal reabsorption. Percentage of viability was calculated as [(number viable pups per litter/total number pups per litter) × 100]. At each time point, the placenta samples were collected for bacteriology, gene expression analysis and blinded histopathological analysis (Extended Data Fig. 6d). When indicated, mice were treated i.p. at days 5, 7 and 9 post-infection with a daily dose of 250 mg per kg body weight of the ER stress inhibitor tauroursodeoxycholate TUDCA (Sigma-Aldrich) or vehicle control.

Real-time PCR. RNA was isolated from BMDMs and mouse tissues using Tri-reagent (Molecular Research Center) according to the instructions of the manufacturer. Reverse transcription was performed on 1 μg of DNase-treated RNA with Taqman reverse transcription reagent (Applied Biosystems). For each real-time reaction, 4 μl of cDNA was used combined with primer pairs for mouse *Actb*, *Il6*, *Hspa5* and *Chop*. Real time transcription-PCR was performed using Sybr green and an ABI 7900 RT-PCR machine (Applied Biosystems). The fold change in mRNA levels was determined using the comparative threshold cycle (C_t) method. Target gene transcription was normalized to the levels of *Actb* mRNA.

Cytokine detection. Cytokine levels in mouse serum and supernatants of infected BMDMs were measured using either a multiplex cytokine/chemokine assay (Bio-Plex 23-plex mouse cytokine assay; Bio-Rad), or via an enzyme-linked immunosorbent assay (IL-6 ELISA; eBioscience), according to the manufacturer's instructions.

LDH release assay. Cytotoxicity was determined by using a LDH release assay in supernatant of BMDMs treated as described above. LDH release assay was performed using a CytoTox 96 Non-Radioactive Cytotoxicity Assay (Promega), following manufacturer's protocol. The percentage of LDH release was calculated as follows:

Percentage of LDH release = 100 × (absorbance reading of treated well – absorbance reading of untreated control)/(absorbance reading of maximum LDH release control – absorbance reading untreated control). The kit-provided lysis buffer was used to achieve complete cell lysis and the supernatant from lysis-buffer-treated cells was used to determine maximum LDH release control.

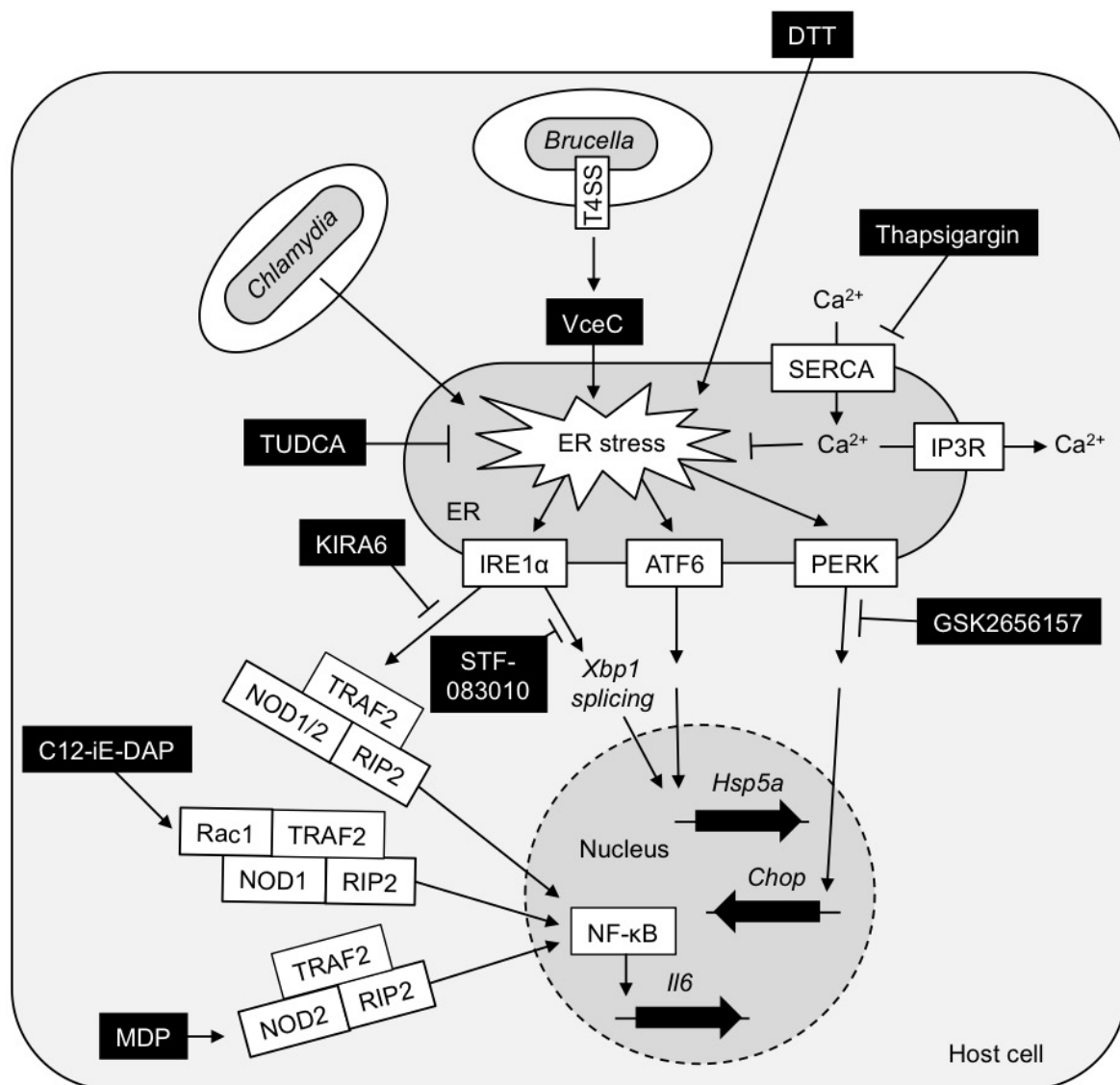
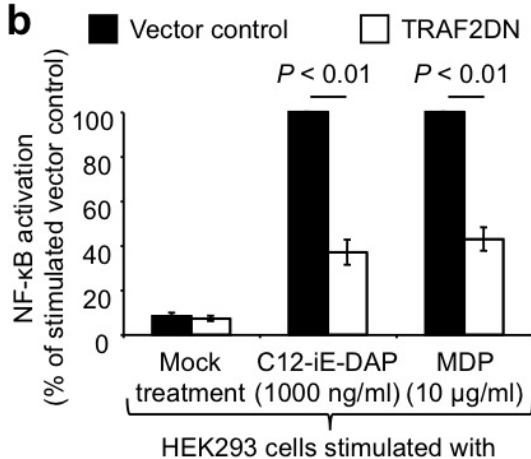
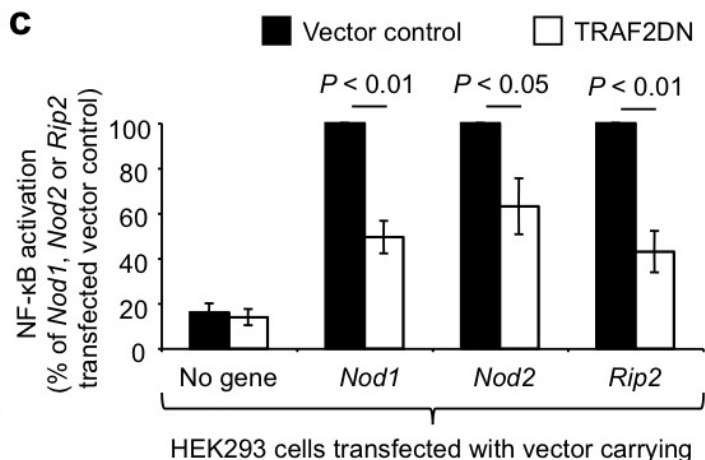
Chlamydia infection (line 545). HeLa 229 cells (ATCC CCL-2.1) were cultured in 96-well tissue culture plates at a concentration of 4 × 10⁴ cells per well in Dulbecco's Modified Eagle Medium (DMEM) (Life Technologies, Grand Island, NY)

supplemented with 10% FBS. HeLa 229 cells were transfected with a total of 125 ng of pCMV-HA-Rip2DN or empty control vector per well. 24 h post-transfection HeLa 229 cells were treated with Dextran to enhance infection efficacy before they were infected with 1.7×10^5 *Chlamydia* bacteria per well. The plates were centrifuged at 2,000 r.p.m. for 60 min at 37 °C, then incubated for 30 min at 37 °C in 5% CO₂. Supernatant was discarded and replaced with DMEM containing 1 µg ml⁻¹ cyclohexine (Sigma Aldrich) and where indicated, 1 µM KIRA6, 10 µM thapsigargin or 10 µg ml⁻¹ MDP, was added to cultures before incubation at 37 °C in 5% CO₂ for 40 h. For gene expression assays, HeLa 229 cells were suspended in Tri-reagent (Molecular Research Center, Cincinnati) and RNA was isolated. Infection efficiency was confirmed in separate plates by staining *Chlamydia*-infected HeLa 229 cells with anti-*Chlamydia* MOMP antibody and counting bacteria under a fluorescent microscope. Four independent assays were performed and the standard error of the mean calculated.

Western blots. BMDMs stimulated where indicated with 10 µM thapsigargin for 24 h were lysed in lysis buffer (4% SDS, 100 mM Tris, 20% glycerol) and 10 µg of protein was analysed by western blot using antibodies raised against rabbit TRAF2 (C192, #4724, Cell Signaling), rabbit HSP90 (E289, #4875, Cell Signaling), mouse SGT1 (ab60728, Abcam) and rabbit α/β-tubulin (#2148, Cell Signaling).

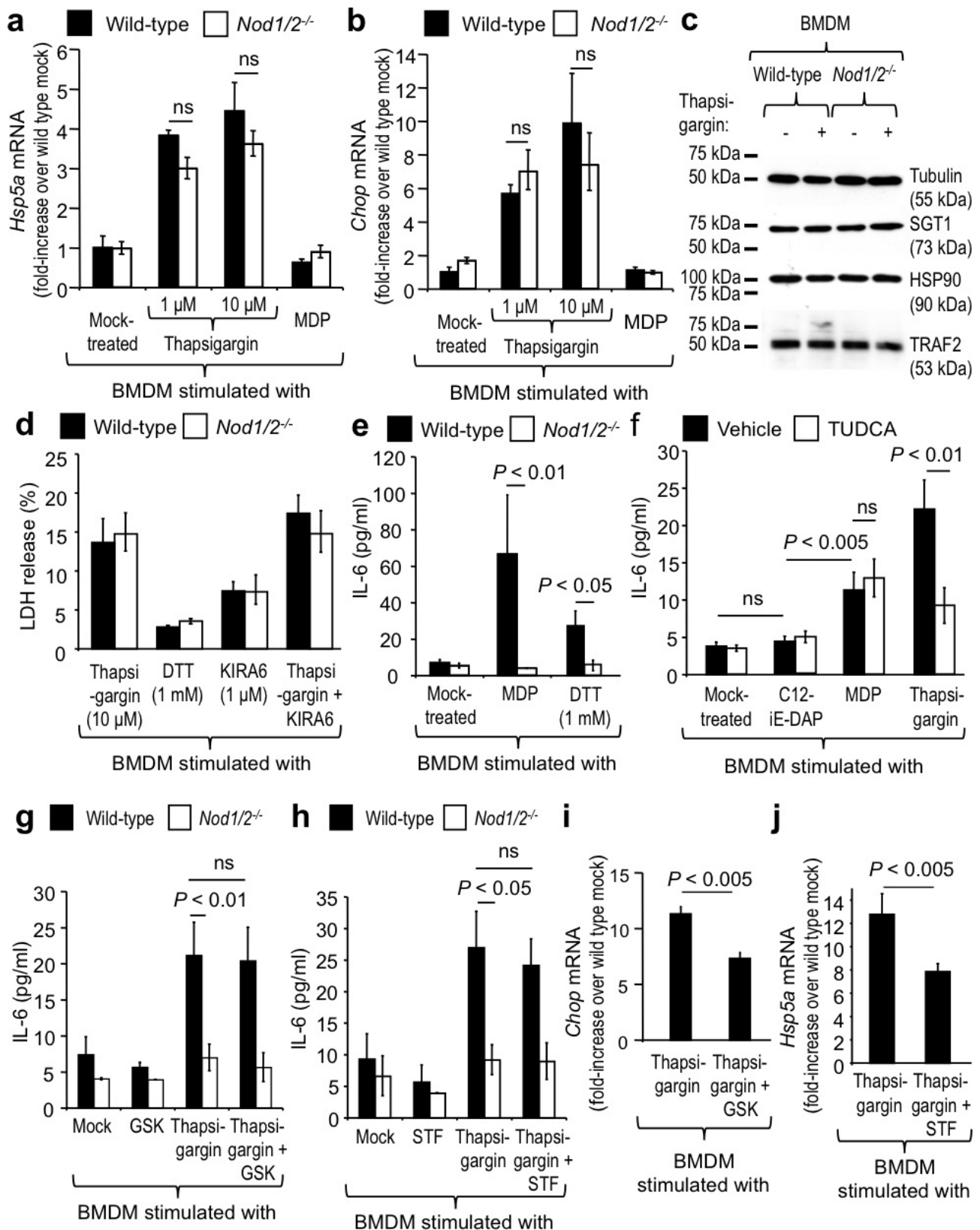
Statistical analysis. For tissue culture experiments, statistical differences were calculated using a paired Student's *t*-test. To determine statistical significance in animal experiments, an unpaired Student's *t*-test was used. To determine statistical significance of differences in total histopathology scores, a Mann-Whitney *U*-test was used. A two-tailed *P* value of <0.05 was considered to be significant.

23. Scidmore, M. A. Cultivation and laboratory maintenance of *Chlamydia trachomatis*. *Curr. Protoc. Microbiol.* **Chapter 11**, Unit 11A.1, (2005).
24. Keestra, A. M. *et al.* A Salmonella virulence factor activates the NOD1/NOD2 signaling pathway. *MBio* **2**, e00266–11 (2011).
25. Keestra, A. M. *et al.* Manipulation of small Rho GTPases is a pathogen-induced process detected by NOD1. *Nature* **496**, 233–237 (2013).
26. Rothe, M., Sarma, V., Dixit, V. M. & Goeddel, D. V. TRAF2-mediated activation of NF-κ B by TNF receptor 2 and CD40. *Science* **269**, 1424–1427 (1995).
27. Rolan, H. G. & Tsois, R. M. Mice lacking components of adaptive immunity show increased *Brucella abortus virB* mutant colonization. *Infect. Immun.* **75**, 2965–2973 (2007).
28. Rolan, H. G. & Tsois, R. M. Inactivation of the type IV secretion system reduces the Th1 polarization of the immune response to *Brucella abortus* infection. *Infect. Immun.* **76**, 3207–3213 (2008).

a**b****c**

Extended Data Figure 1 | Schematic of ER stress and NOD1/2 signalling. **a**, Model of how ER stress induces a NOD1/2-dependent pro-inflammatory response through a TUDCA/KIRA6-sensitive pathway, which differs from the TUDCA/KIRA6-resistant pathways induced by bacterial peptidoglycan fragments (MDP or C12-iE-DAP). **b**, NF-κB activation induced by ectopic expression of VceC in HEK293 cells

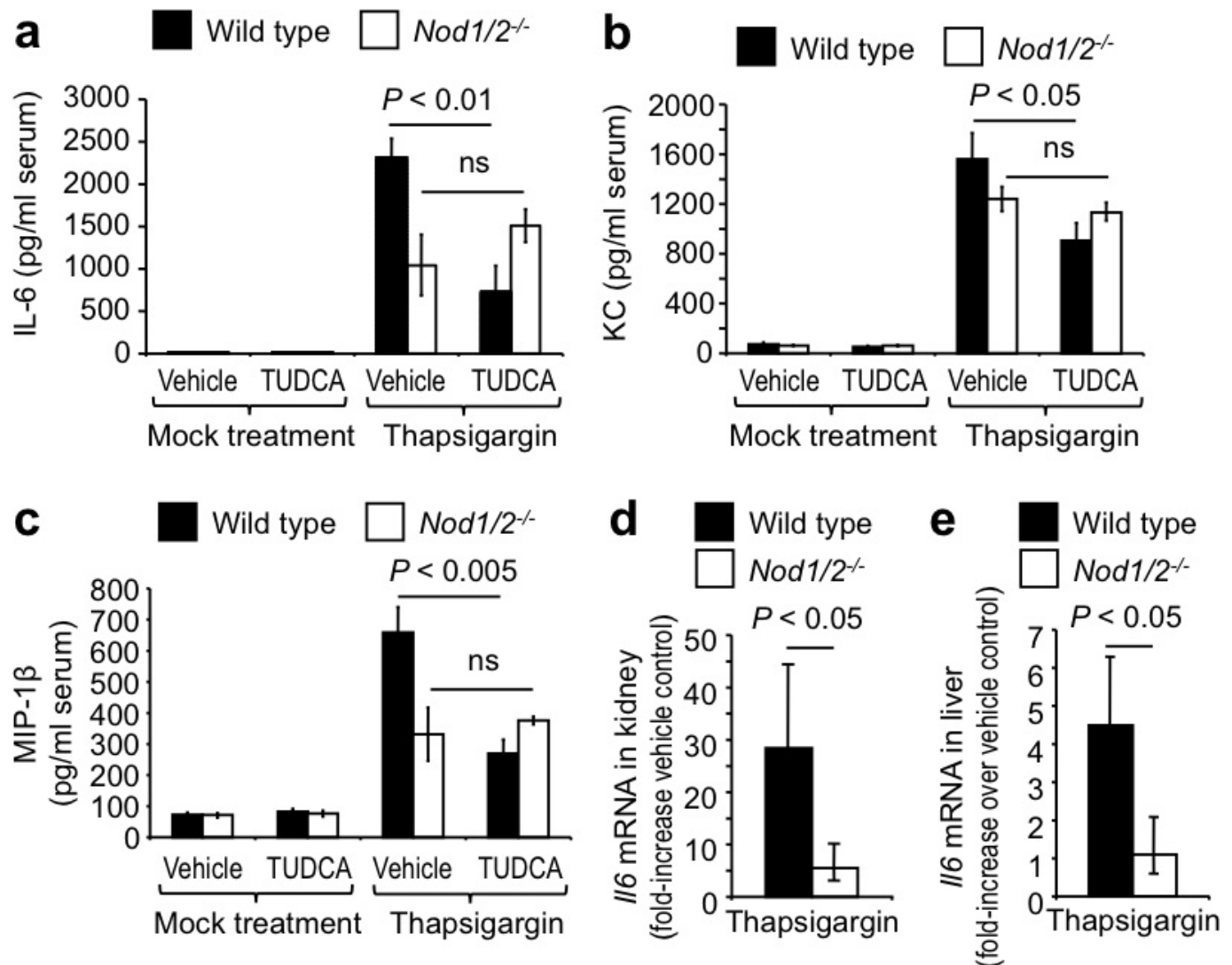
transfected with a dominant negative form of TRAF2 or a vector control. **c**, NF-κB activation mediated by expression-induced auto-activation of NOD1, NOD2 or RIP2 in HEK293 cells that were transfected with a dominant negative form of TRAF2 or a vector control. Data are expressed as mean luciferase activity \pm s.e.m. from five independent experiments.



Extended Data Figure 2 | See next page for caption.

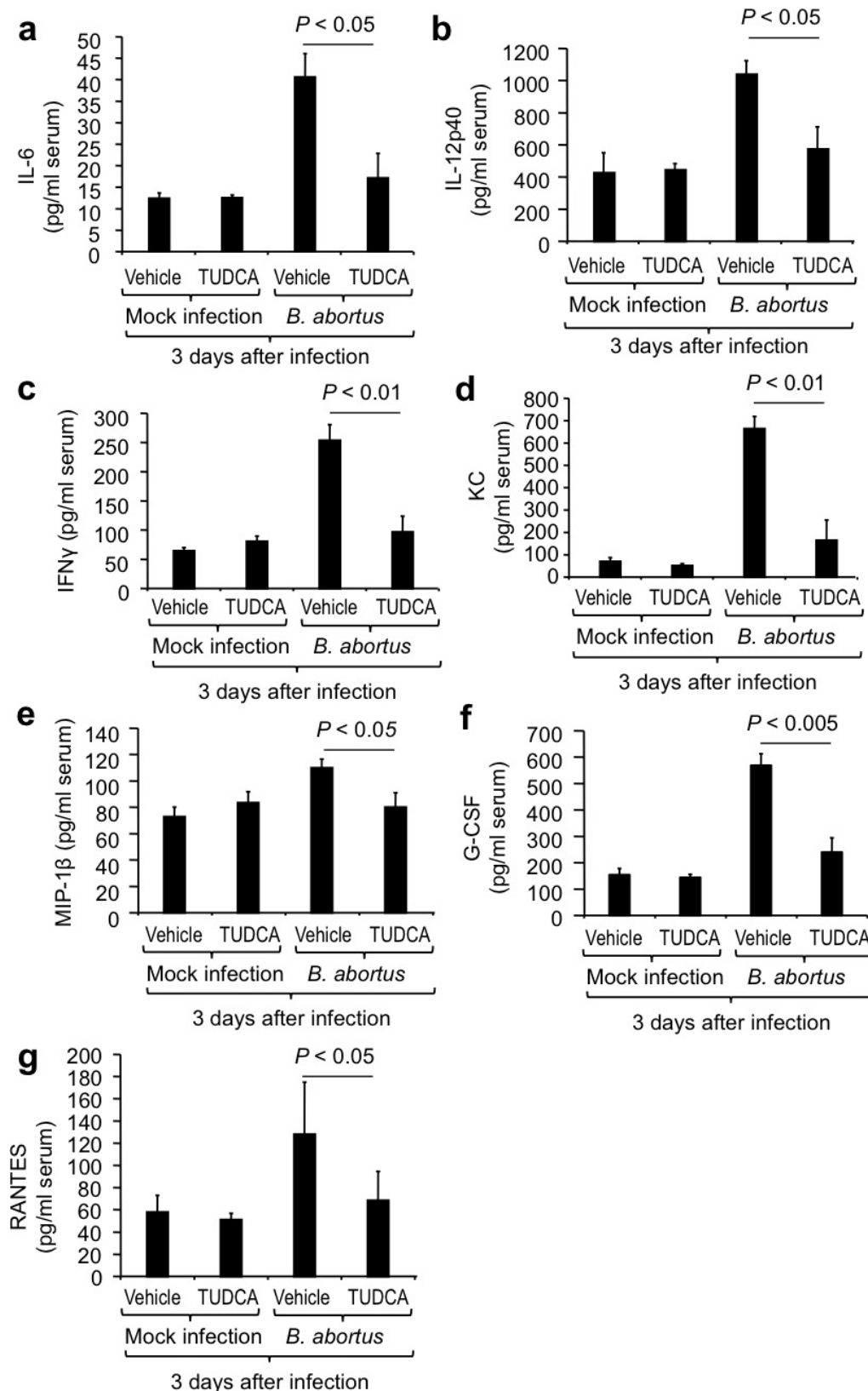
Extended Data Figure 2 | Only the pro-inflammatory arm of the UPR requires NOD1 and NOD2. **a, b**, BMDMs from *Nod1/2^{-/-}* mice and wild-type littermates were stimulated with thapsigargin or MDP, and mRNA abundance for *Hspa5* (**a**) and *Chop* (**b**) was quantified ($n = 4$). **c**, Expression of SGT1, HSP90 and TRAF2 was detected by western blot in lysates of thapsigargin-stimulated BMDMs from wild-type mice and *Nod1/2^{-/-}* mice ($n = 3$ mice of each genotype). Detection of tubulin served as a loading control. A representative image for BMDMs from one wild-type and one *Nod1/2^{-/-}* animal is shown. **d**, LDH release induced by treatment of BMDMs from wild-type mice and *Nod1/2^{-/-}* mice ($n = 5$) with thapsigargin, DTT or KIRA6. **e**, Stimulation with MDP or

DTT induced IL-6 production in BMDMs from C57BL/6 mice (wild type) but not in BMDMs from *Nod1/2^{-/-}* mice ($n = 8$). **f**, IL-6 secretion induced by thapsigargin, but not by the canonical NOD2 ligand MDP, was significantly inhibited by ER stress inhibitor TUDCA in BMDMs ($n = 8$). BMDMs did not respond to stimulation with a canonical NOD1 ligand (C12-iE-DAP). **g–k**, BMDMs from wild-type mice and *Nod1/2^{-/-}* mice ($n = 4$) were treated with the PERK inhibitor GSK2656157 (GSK) (**g, i**) or the IRE1 α RNase inhibitor STF-083010 (STF) (**h, k**) and IL-6 synthesis measured by ELISA (**g, h**) or mRNA analysed by real-time PCR (**i, j**). Data are presented as mean \pm s.e.m. n represents the number of independent assays (biologic replicates) performed for each experiment.



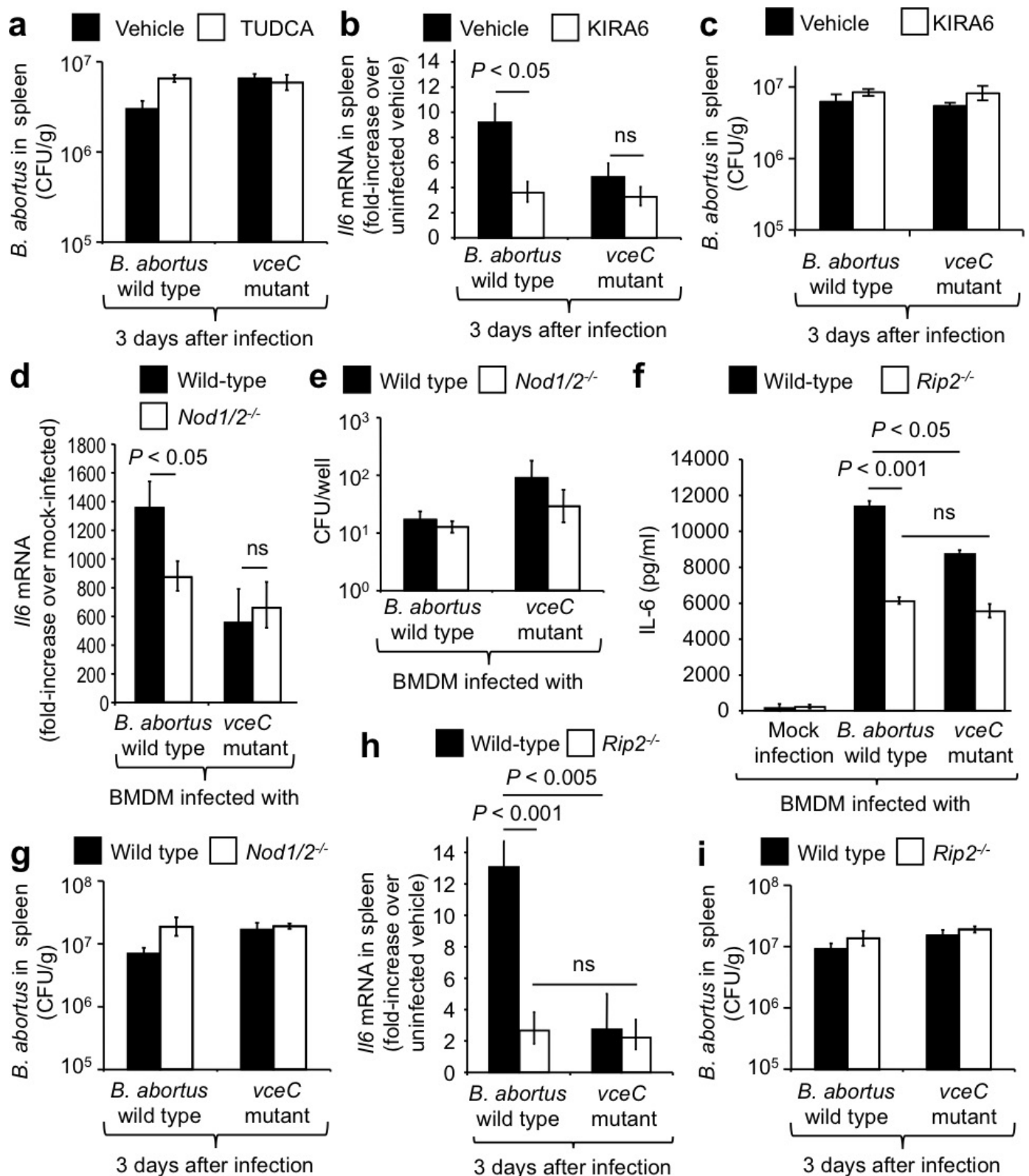
Extended Data Figure 3 | Proinflammatory responses induced by thapsigargin are NOD1/2-dependent. a–c, Groups ($n=5$) of wild-type mice and *Nod1/2*^{-/-} mice were treated with thapsigargin and received either vehicle control or TUDCA. Synthesis of IL-6 (a), KC (b) and MIP-1β (c) in the serum was determined using a Bio-plex cytokine assay. d, e,

Wild-type (C57BL/6) mice and *Nod1/2*^{-/-} mice ($n=4$) were treated with thapsigargin and transcript levels of *Il6* determined by quantitative real-time PCR. Data are expressed as fold-increases over vehicle control-treated animals. Data are presented as mean \pm s.e.m. n represents the number of independent assays (biologic replicates) performed for each experiment.



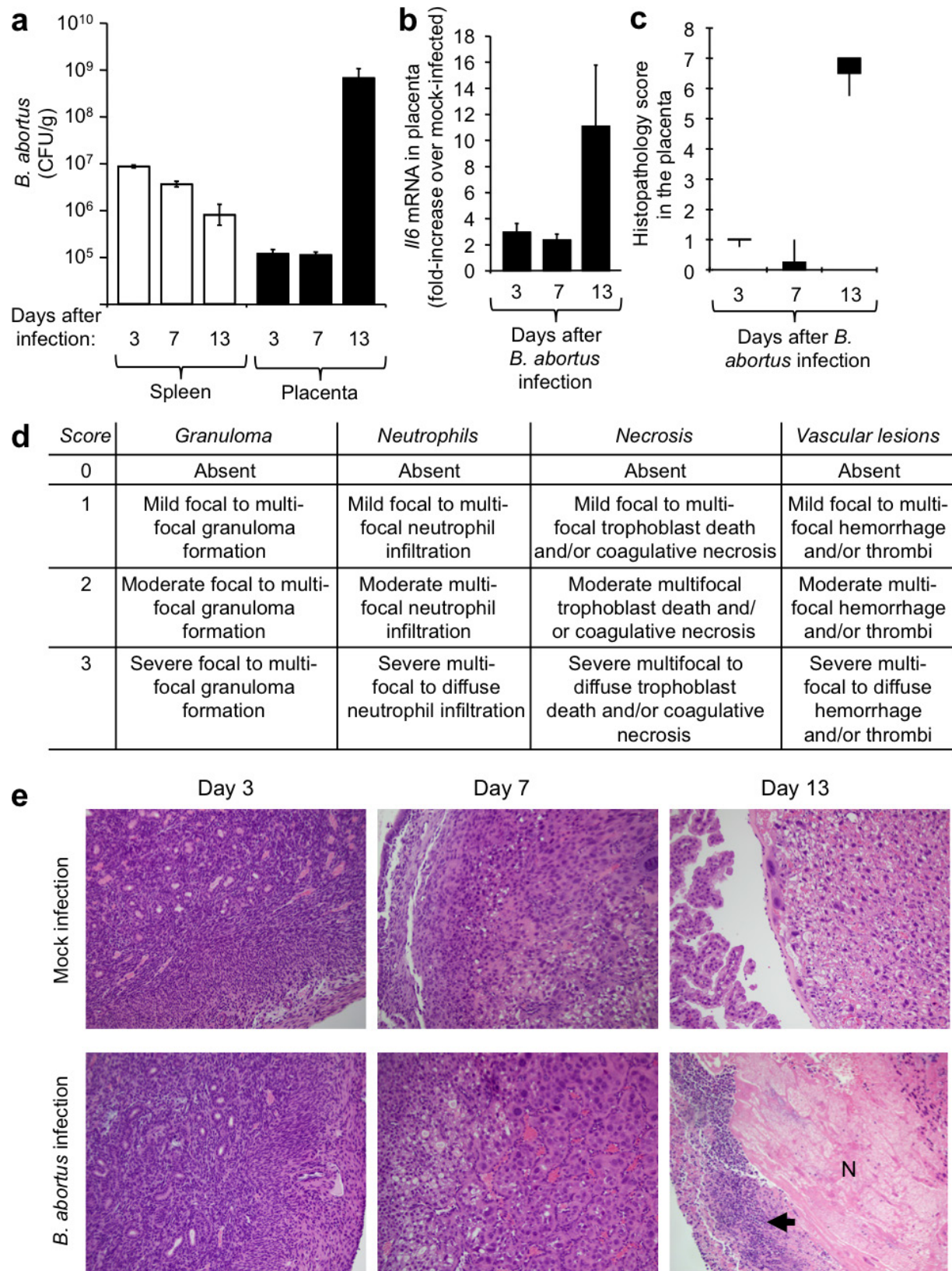
Extended Data Figure 4 | *B. abortus*-induced inflammatory responses in mice are blunted by TUDCA treatment. a–g. Mice ($n \geq 4$) were mock infected or infected with the *B. abortus* wild type and were treated with TUDCA or vehicle control. Three days after infection, circulating levels

of IL-6 (a), IL-12p40 (b), IFN γ (c), KC (d), MIP-1 β (e), G-CSF (f) and RANTES (g) were profiled in serum using a Bio-Plex cytokine assay. Data are presented as mean \pm s.e.m.

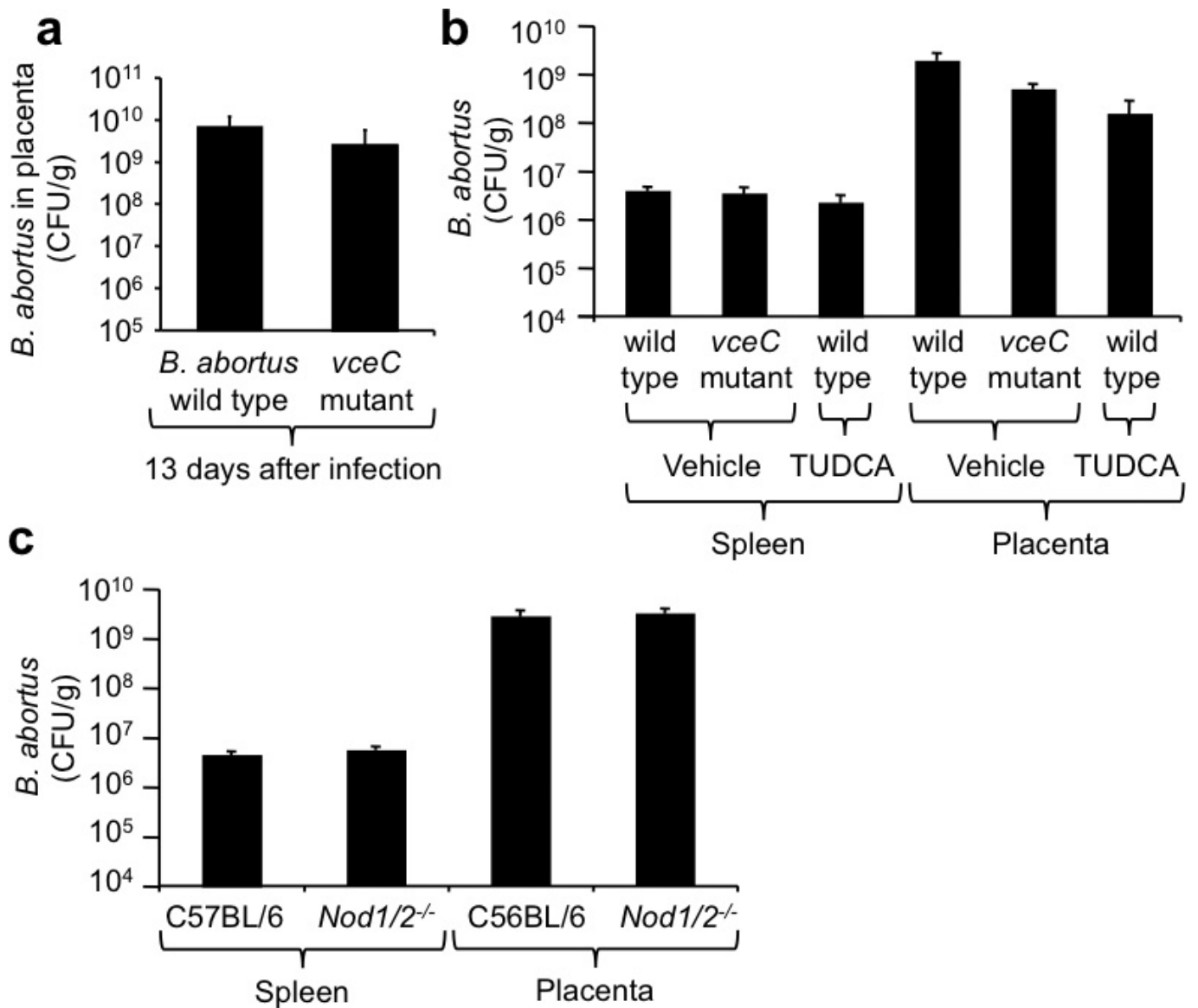


Extended Data Figure 5 | Bacterial burden and host responses during infection with *B. abortus*. a, c, e, g, i, Bacterial burden in the spleen and in BMDMs of wild-type and *Nod1/2^{-/-}* mice (a, c, e, g) or *Rip2^{-/-}* mice (i). No statistically significant differences in colony-forming units (CFU) recovered from the spleen (a, c, g, i) or from BMDMs (e) of wild-type and *Nod1/2^{-/-}* or *Rip2^{-/-}* mice ($n \geq 4$) infected with *B. abortus* wild type or the *vceC* mutant were observed. b, d, f, h, Host responses elicited during *B. abortus* infection. b, Groups of mice ($n = 5$) were infected with the

indicated *B. abortus* strains and treated with KIRA6. d, f, BMDMs from wild-type mice and *Nod1/2^{-/-}* mice (d) or wild-type mice and *Rip2^{-/-}* mice ($n \geq 4$) (f) were infected with the indicated *B. abortus* strains. h, Groups ($n = 5$) of wild-type mice and *Rip2^{-/-}* mice were infected with the indicated *B. abortus* strains. *Il6* mRNA levels were determined by quantitative real-time PCR (b, d, h). IL-6 synthesis was determined by ELISA (f). Data are presented as mean \pm s.e.m.

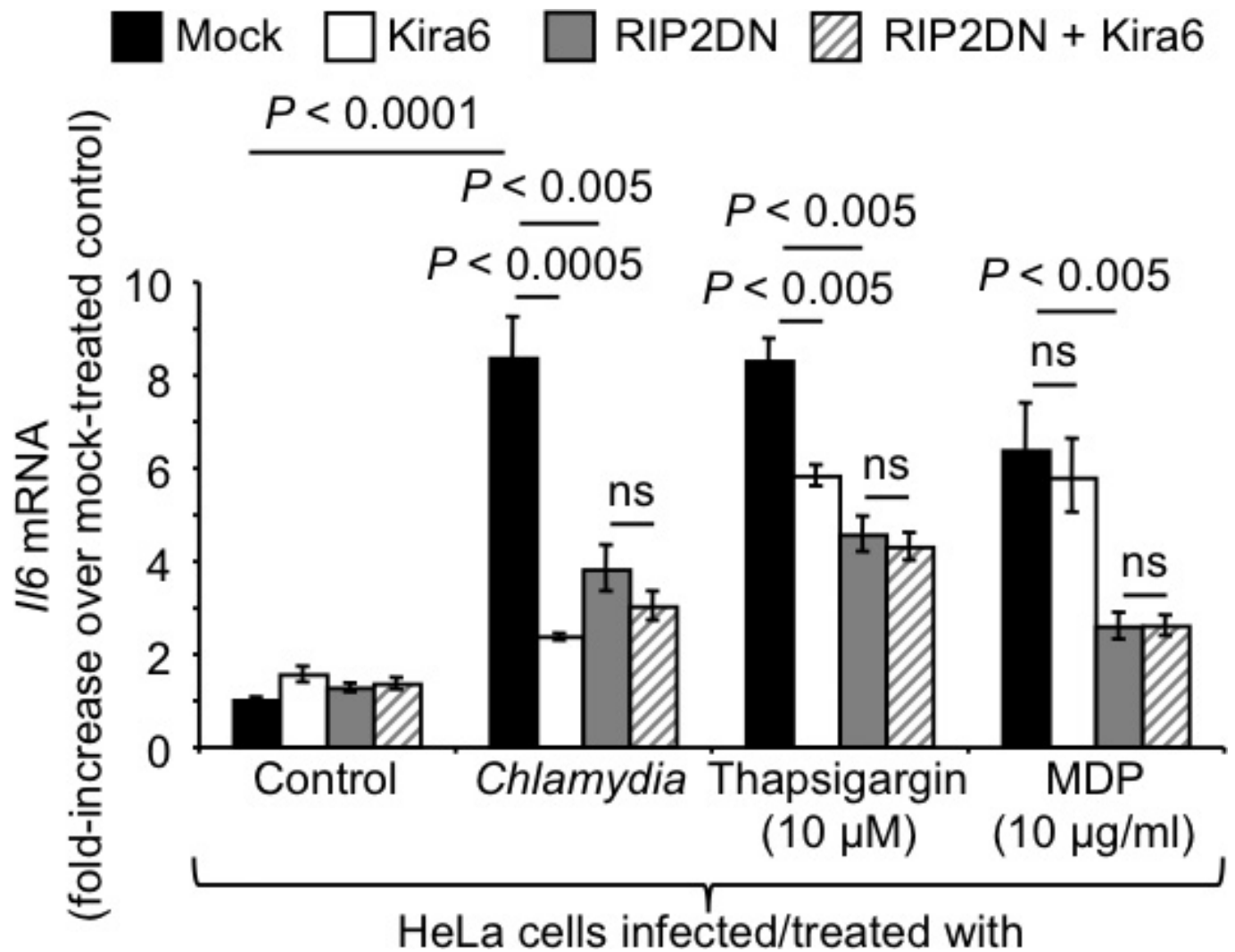


Extended Data Figure 6 | The *B. abortus* placentitis model. **a**, Bacterial numbers of wild-type *B. abortus* (strain 2308) recovered from in the spleen and placenta ($n = 5$ mice per group). **b**, **c**, *Il6* mRNA expression (**b**) and total histopathology scores (**c**) in the placenta of mice at days 3, 7 and 13 after infection with *B. abortus*. **d**, Scoring criteria for blinded evaluation of haematoxylin and eosin (H&E)-stained sections from the placenta. **e**, Representative images of the histopathology observed in the placenta of *B. abortus* infected mice at days 3, 7 and 13 after infection. Arrow, neutrophil infiltration; N, necrosis.



Extended Data Figure 7 | Bacterial burden in the spleen and placenta of wild-type and *Nod1/2*^{-/-} mice. a–c, No statistically significant differences in colony-forming units in the spleen and placenta of

wild-type and *Nod1/2*^{-/-} mice infected with *B. abortus* wild type or the *vceC* mutant at 13 days post-infection were observed. Data are presented as mean \pm s.e.m. ($n = 5$ mice per group).



Extended Data Figure 8 | *IL6* expression induced by *Chlamydia muridarum*. HeLa cells were stimulated with MDP, thapsigargin or infected with *Chlamydia muridarum* and treated with KIRA6 or

transfected with RIP2DN (dominant negative form of RIP2). Expression of *IL6* was determined by quantitative real-time PCR. Data are presented as mean \pm s.e.m. from 4 independently performed assays.

USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites

Byung-Hoon Lee¹, Ying Lu², Miguel A. Prado¹, Yuan Shi^{1†}, Geng Tian¹, Shuangwu Sun^{1,3}, Suzanne Elsasser¹, Steven P. Gygi¹, Randall W. King¹ & Daniel Finley¹

USP14 is a major regulator of the proteasome and one of three proteasome-associated deubiquitinating enzymes^{1–9}. Its effects on protein turnover are substrate-specific, for unknown reasons. We report that USP14 shows a marked preference for ubiquitin–cyclin B conjugates that carry more than one ubiquitin modification or chain. This specificity is conserved from yeast to humans and is independent of chain linkage type. USP14 has been thought to cleave single ubiquitin groups from the distal tip of a chain, but we find that it removes chains from cyclin B en bloc, proceeding until a single chain remains. The suppression of degradation by USP14's catalytic activity reflects its capacity to act on a millisecond time scale, before the proteasome can initiate degradation of the substrate. In addition, single-molecule studies showed that the dwell time of ubiquitin conjugates at the proteasome was reduced by USP14-dependent deubiquitination. In summary, the specificity of the proteasome can be regulated by rapid ubiquitin chain removal, which resolves substrates based on a novel aspect of ubiquitin conjugate architecture.

The proteasome, a ~2.5-MDa molecular machine, is the focal point for global regulation of ubiquitin pathway output¹⁰. Among its major regulators is a deubiquitinating enzyme known in mammals as USP14 (in yeast, Ubp6). USP14 negatively regulates proteasome activity by ubiquitin chain disassembly as well as by a noncatalytic mechanism^{1–6}. USP14 inhibitors produce selective effects on the turnover of proteasome substrates^{3,8}, suggesting that the rate at which USP14 disassembles proteasome-bound ubiquitin chains may depend on the nature of the substrate. We test this hypothesis in the present work.

USP14 is thought to progressively trim monoubiquitin groups from the substrate-distal tip of a chain⁹ (Fig. 1a). To assess this model, we employed a canonical substrate of USP14, ubiquitinated cyclin B1 (refs 1, 3). When cyclin B1 is modified by the ubiquitin ligase anaphase-promoting-complex/cyclosome (APC/C) in the presence of E2 enzyme UbcH10, the resulting conjugates typically carry multiple short ubiquitin chains of varied length^{11,12}. We refer to these as supernumerary chains, insofar as only one chain is in principle required for substrate degradation. The flexible, 88-residue amino-terminal element of cyclin B1 (NCB1) is enriched in lysine residues, nine of which are competent for ubiquitination¹¹. The N-terminal element is necessary for cyclin B1 degradation, and contains an APC/C recognition motif, the destruction box^{11–13}.

Ubiquitinated NCB1 was rapidly degraded by proteasomes lacking USP14, but in the presence of USP14 deubiquitinated species were observed. These were not fully deubiquitinated, but rather carried 2–4 ubiquitin groups and were resistant to both degradation and further deubiquitination (Fig. 1b). This strong stop to USP14-dependent processing might be assumed to reflect that these conjugates retain too few ubiquitin moieties for efficient proteasome binding. However, studies below show that USP14 activity depends on the architecture

of ubiquitin chains on the substrate and not simply substrate binding to the proteasome. The behaviour of NCB1 was comparable to that of full-length cyclin B1 (data not shown)^{3,12}.

We used the experimental system described above to test whether the architecture of ubiquitin chains on NCB1 was relevant to its rapid

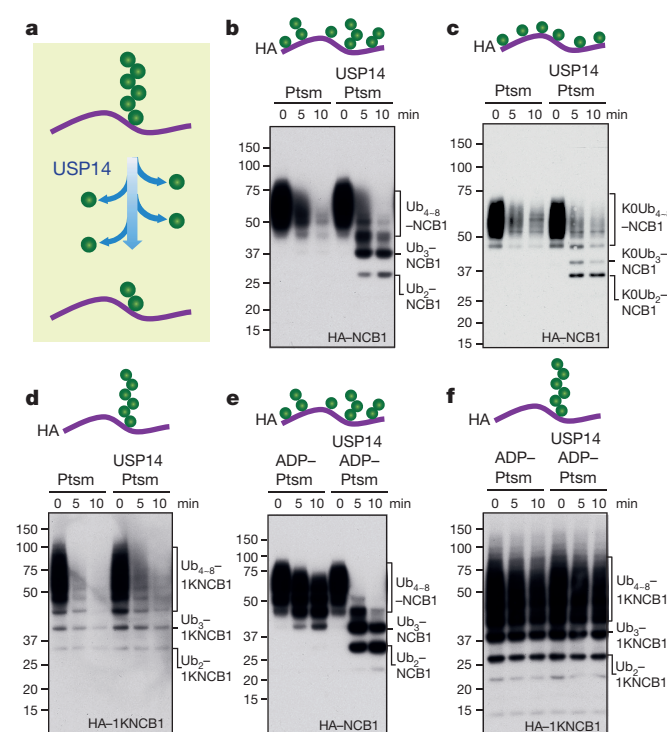


Figure 1 | USP14 cleaves supernumerary ubiquitin chains from substrates. **a**, Canonical model of USP14 activity. USP14 is thought to shorten ubiquitin chains progressively from their distal tip. **b–f**, *In vitro* degradation and deconjugation assays with polyubiquitinated N-terminal fragment of cyclin B1 (Ub_n-NCB1; haemagglutinin (HA) tagged). Cartoons show idealized representations of the expected distribution of ubiquitin moieties throughout. **b**, Human proteasome (Psm; 4 nM) was incubated with Ub_n-NCB1 (~110 nM final) generated by UbcH10. Where indicated, USP14 (80 nM) was added. **c**, As in **b**, except using polyubiquitinated NCB1 generated with lysine-less Ub. **d**, As in **b**, except using polyubiquitinated conjugates on K64-only NCB1 (1KNCB1, unless otherwise noted), generated by UbcH10 and Ube2S. **e**, Ub_n-NCB1 (~110 nM) was incubated with hexokinase-treated ADP-proteasome (ADP-Psm) in the presence or absence of USP14 (80 nM). **f**, As in **e**, except using polyubiquitinated K64-only NCB1. Samples were analysed by SDS-PAGE/immunoblotting for HA. For gel source data, see Supplementary Information.

¹Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. ³Life Sciences Institute, Zhejiang University, Hangzhou, 310058, China. [†]Present address: Department of Molecular and Clinical Pharmacology, Factor 10-638, 650 Charles E. Young Drive South, University of Los Angeles, Los Angeles, California 90095, USA.

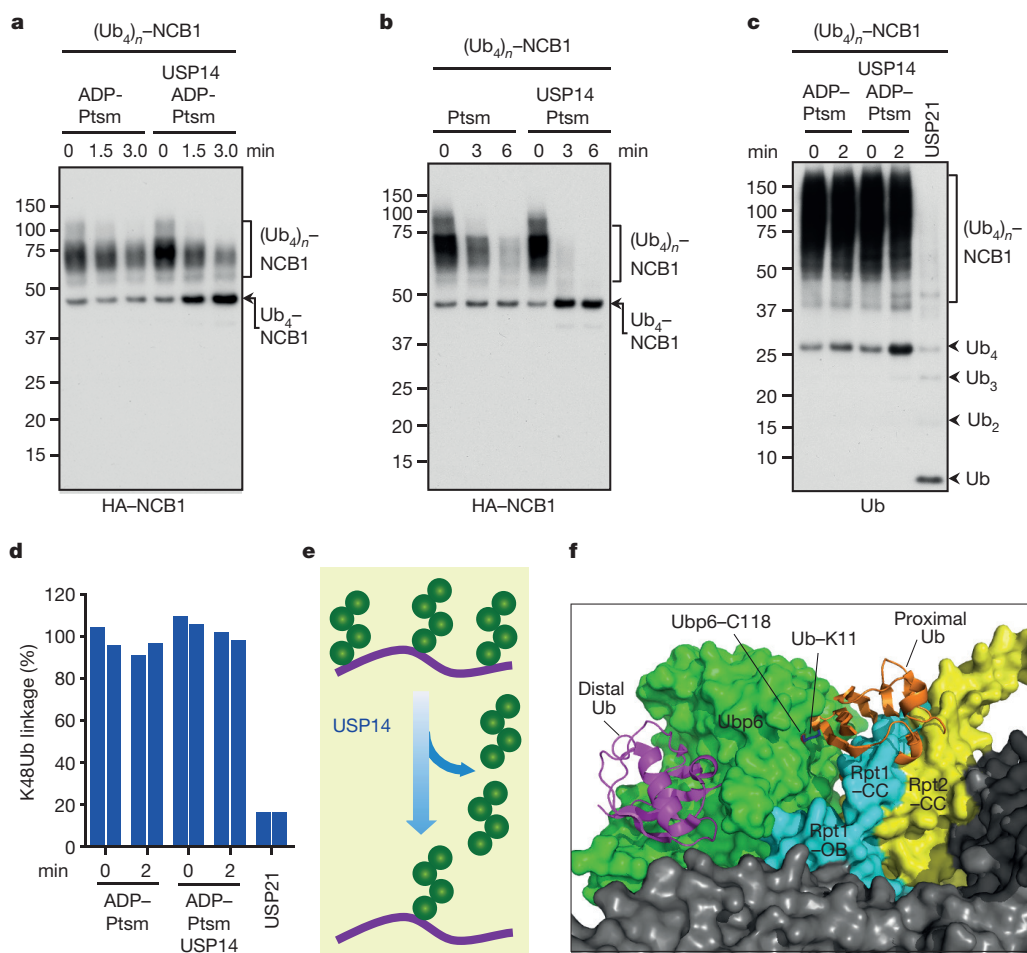


Figure 2 | USP14 cleaves arrayed tetraubiquitin chains proximally, stopping at the last chain. **a**, K48-linked Ub₄ conjugates of NCB1 were prepared, resulting in one or more tetraubiquitin modifications. ADP-Ptsm (3 nM) was incubated with multi-K48Ub₄-NCB1 (~90 nM) in the presence or absence of USP14 (60 nM). Arrow indicates K48Ub₄-NCB1 species modified with a single chain and derived by USP14 cleavage. **b**, Assays as in **a**, except in the presence of ATP. **c**, K48Ub₄-multichain conjugate of NCB1 was purified, and assays were performed as in **a**, except that a control was added in which the pan-specific deubiquitinating enzyme USP21 (0.5 μM) was used (10 min incubation). Arrowheads indicate free Ub₄ chains released by USP14 (K48Ub₄) and products of USP21 activity such as free Ub. Samples were analysed by SDS-PAGE/immunoblotting for HA (**a** and **b**) or Ub (**c**). **d**, Tandem mass tag (TMT)

labelling and quantitative mass spectrometry. Purified K48Ub₄-multi-chain NCB1 conjugates were assayed in biological duplicates for each condition as in **a**. The reaction was then quenched and subjected to trypsin digestion, TMT labelling, and tMS² analysis in a Q-Exactive for quantification of K48-linked Ub peptides. **e**, Proposed en bloc mode of deubiquitination by USP14. **f**, Modelling of free ubiquitin chain cleavage by proteasome-activated Ubp6. Access of a K11-linked diubiquitin chain to the active site of Ubp6 seems to be occluded. Structures were obtained from the PDB database: Ubp6 and proteasome (5A5B)⁵ and Ub (1UBQ). C118 contains the active site nucleophile. Note that occlusion was similarly observed for modeled ubiquitin dimer linked via either K48 or K63 (Extended Data Fig. 9).

deubiquitination. Cyclin B1 is modified by APC/C and UbcH10 to form mixed-linkage chains¹¹ primarily via residues K11, K48, and K63. Thus, a lysine-free mutant of ubiquitin¹² will modify NCB1 via multiple mono-ubiquitination events (Fig. 1c). USP14 deubiquitinated this form of NCB1 at a rate comparable to that of wild-type conjugates, indicating that USP14 does not act obligatorily on ubiquitin-ubiquitin linkages (Fig. 1c).

In a complementary experiment, we eliminated lysines from NCB1 rather than ubiquitin, leaving behind only K64 of NCB1 to allow for the synthesis of a single ubiquitin chain on NCB1. This mutant is competent for degradation (Fig. 1d)¹². Although single-chain conjugates are canonical proteasome substrates¹⁴, USP14 showed no detectable activity on this substrate (Fig. 1d). The lack of deubiquitinating activity on the single-chain substrate was accompanied by a failure to inhibit its degradation. Similar results were obtained with NCB1 ubiquitinated with Ubc4 in place of UbcH10 (Extended Data Fig. 1a, b).

A caveat to the experiment above is that, because NCB1 is degraded quickly by the proteasome, USP14 has little time to act. However, the

proteasome must be present in the assay because it is required to activate USP14. When we quenched substrate degradation by replacing ATP with ADP (Extended Data Fig. 2), USP14-dependent deubiquitination of WT NCB1 conjugates was preserved (Fig. 1e). In contrast, no deubiquitination was detected with single-chain conjugates (Fig. 1f). Thus, USP14 shows marked specificity for NCB1 conjugates bearing multiple ubiquitin chains, whereas the proteasome does not effectively distinguish between these two classes of substrate in the absence of USP14.

We next considered whether the resistance of single-chain conjugates to USP14 was idiosyncratic to the particular modified lysine in NCB1. On the contrary, conjugates generated using a K36-only form of NCB1 behaved equivalently to the K64-anchored conjugates (Extended Data Fig. 1c). The supernumerary chain effect was seen with the substrate PY-Sic1 as well as NCB1 (Extended Data Fig. 1d, e). We identified a third preferred substrate of USP14, securin, which is likewise modified at multiple lysines (Extended Data Fig. 1f)^{13,15}. Ubp6, the *Saccharomyces cerevisiae* orthologue of USP14, showed the same preference for supernumerary chains, indicating that this property is evolutionarily conserved (Extended Data Fig. 3).

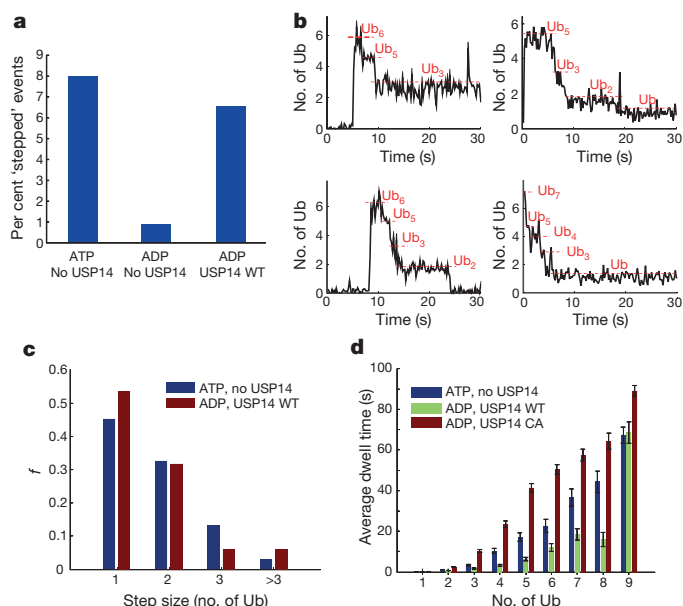


Figure 3 | Single-molecule analysis of deubiquitination by USP14. Single-molecule assays were performed on immobilized human proteasome from which USP14 had been removed. Securin conjugates were generated using fluorescently labelled ubiquitin and added to the proteasome under conditions as indicated. **a**, Percent of stepped traces under conditions measuring RPN11 (ATP, no USP14) or USP14 (ADP, USP14-WT) activities. N (ATP, no USP14) = 1,574; N (ADP, no USP14) = 578; N (ADP, USP14-WT) = 733. **b**, Examples of step-trace segmentation graphs from a single-molecule analysis on polyubiquitinated securin conjugates with USP14-reconstituted human proteasome in the presence of ADP (See Extended Data Fig. 10 for RPN11 step traces). **c**, Step size distribution of RPN11-mediated (ATP, no USP14, N = 356) or USP14-mediated (ADP, USP14-WT, N = 129) deubiquitination events as a function of the number of ubiquitins. **d**, Average dwell time of securin conjugates on the proteasome in the presence or absence of USP14. ATP was added for the no-USP14 condition. Dwell time was plotted as a function of the multiplicity of ubiquitin groups on securin. ADP denotes the condition of ADP plus *o*-phenanthroline to suppress RPN11 activity. USP14-CA is a catalytically inactive mutant of USP14 (Cys114Ala). Error bars indicate s.d. of the mean from three independent experiments. N (ATP, no USP14) = 948; 1,097; 691; 523; 382; 301; 196; 133; 242 (for Ub to Ub₉, respectively); N (ADP, USP14-WT) = 1,854; 1,789; 1,129; 815; 536; 274; 164; 86; 96; N (ADP, USP14-CA) = 598; 764; 640; 574; 507; 437; 289; 173; 245.

It has been hypothesized that USP14 or Ubp6 may preferentially cleave K63-linked chains^{16,17}. This model could account for our data if single-chain NCB1 conjugates were devoid of K63 linkages. To test this model, we prepared single-chain NCB1 with homogeneous K63 linkages. These conjugates were also cleaved slowly if at all by USP14 (Extended Data Fig. 4), regardless of length.

One explanation for the specificity of USP14 for substrates modified at multiple sites is that cleavage of one chain is promoted allosterically by binding of a second chain to USP14 or another site on the proteasome. However, adding free chains over a wide concentration range failed to stimulate deubiquitination of single-chain NCB1 (Extended Data Fig. 5). Another measure that failed to stimulate removal of single-chain chains was phosphomimetic substitution of Ser432, a site subject to modification by Akt¹⁸ (Extended Data Fig. 6). If singleton chains are as a rule poor substrates for USP14, then all free ubiquitin chains—those not attached to an acceptor protein—must be poor USP14 substrates, regardless of length or linkage type. Extended Data Fig. 7 shows that free chains of diverse length and linkage type are all cleaved minimally by proteasome-activated USP14, even upon long incubation.

By ubiquitinating NCB1 with preformed tetraubiquitin chains, we prepared conjugates carrying multiple chains of homogeneous linkage type and length. When these conjugates were incubated with USP14,

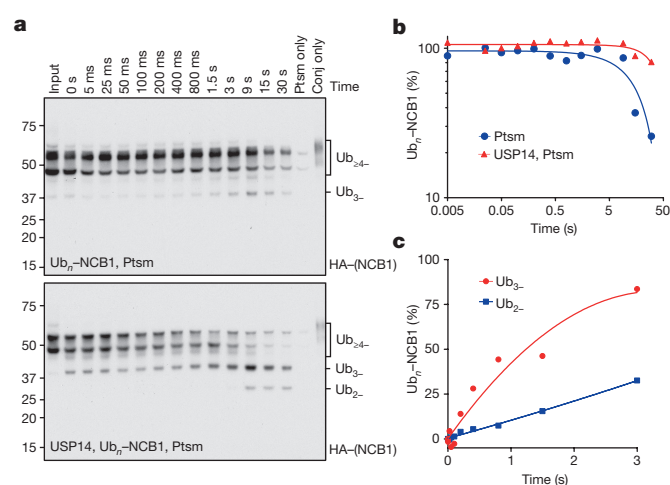


Figure 4 | Kinetic competition between USP14 and the proteasome.

a, Proteasome (1.2 μ M) in the presence or absence of USP14 (17 μ M) was mixed with Ub_n-NCB1 (~0.47 μ M) in a quench flow device. Top, degradation of NCB1 conjugates is detectable by approximately 15 s. Bottom, deubiquitination of Ub₄₋₆-NCB1 can be observed by 100 ms or earlier as Ub₂₋₃-NCB1 species are generated. **b**, Quantification of Ub_n-NCB1 degradation from **a**. **c**, Quantification of Ub₃₋ or Ub₂₋-NCB1 species generated by USP14-mediated deubiquitination. USP14-derived reaction products can appear as early as hundreds of ms. Collected samples were analysed by SDS-PAGE/immunoblotting, using antibody to HA. Note that, for **b**, axes are logarithmic in scale.

proteasomes, and either ATP or ADP, a limit digest product was formed consisting of NCB1 carrying four ubiquitin groups (Fig. 2a, b). Thus, USP14 seemed to completely remove each supernumerary chain, and stop when a single chain remained. Accordingly, ubiquitin was released from this substrate in the form of tetraubiquitin chains (Fig. 2c). Using mass spectrometry, we confirmed that K48 linkages were not consumed in this reaction (Fig. 2d). Studies with other linkage types (Extended Data Fig. 8) further indicated that USP14 removes chains preferentially en bloc (Fig. 2e), rather than trimming them from the distal tip, as previously thought.

Why does USP14 discriminate against cleavage within chains? The presumptive active form of Ubp6 sits with its catalytic domain directly contacting the OB domain of subunit Rpt1 (refs 5,6). Modelling studies with diubiquitin as substrate indicated that, regardless of chain linkage type, the proximal ubiquitin may have restricted access to the active site of the enzyme (Fig. 2f, Extended Data Fig. 9). The coiled-coils of Rpt1 and Rpt2 are responsible for this occlusion (Fig. 2f). This effect may account for the dominant en bloc cleavage mode of Ubp6 and USP14. Cleavage within ubiquitin chains by free forms of USP14 and Ubp6, as observed by many researchers, may reflect that free USP14 is not subject to occlusion from proteasome subunits.

In a recent study, we established a single-molecule TIRF (total internal reflection fluorescence) microscopy method to monitor deubiquitination events mediated by the en bloc zinc-dependent activity of RPN11^{19,20}, using USP14-free proteasomes¹⁵. Deubiquitination was identified as a stepwise reduction of signals from fluorescently-labelled ubiquitin on substrates (Fig. 3a, Extended Data Fig. 10). To study USP14-mediated deubiquitination, we inhibited RPN11 activity (as well as substrate degradation) using ADP and a zinc chelator¹⁵. ‘Stepped’ events representing deubiquitination increased sevenfold upon USP14 addition (Fig. 3a). Single-molecule traces with USP14-reconstituted proteasome in the presence of ADP showed remarkable similarity to those produced by RPN11 (Fig. 3b, Extended Data Fig. 10). The step size of ubiquitin removal by RPN11 and USP14 followed a similar distribution, supporting en bloc cleavage by USP14 (Fig. 3c).

USP14-mediated deubiquitination reduced the dwell time of ubiquitin conjugates on the proteasome, suggesting that rapid deubiquitination can suppress degradation of the substrate by contracting the duration of its encounter with the proteasome (Fig. 3d). This dwell time effect

suggests that, to suppress degradation, USP14 must remove chains at least as rapidly as the substrate is degraded in the absence of USP14. To test this scenario we used quench-flow methods, which allow reactions to be followed on a millisecond time scale. An excess of proteasome over substrate was employed to approximate single-turnover conditions.

USP14-free proteasomes degraded NCB1 conjugates efficiently, with most degradation occurring between 9 and 30 s (Fig. 4a, b). USP14 delayed degradation (Fig. 4b), as high molecular weight conjugates were instead converted into partially deubiquitinated species. The production of lower molecular weight species through deubiquitination was initiated before degradation was evident with USP14-free proteasomes (Fig. 4a–c). Thus, USP14 is sufficiently fast that deubiquitination can be accomplished before the proteasome would otherwise degrade NCB1. Similar results were obtained using full-length cyclin B1 (data not shown), and we confirmed that deubiquitination was mediated by proteasome-bound rather than free USP14 (Extended Data Fig. 6b).

The ability of USP14 to prevail over the proteasome in a kinetic competition may explain the suppression of protein degradation by USP14's deubiquitinating activity³. Thus, USP14-loaded proteasomes may discriminate against substrates bearing multiple ubiquitin chains, whereas USP14-free proteasomes do not. Because USP14 and Ubp6 are modulated by growth factors and stress conditions^{2,18}, the ability of the proteasome to discriminate between single-chain and multi-chain substrates is likely to be under biological control.

We expect that the supernumerary chain rule should bias USP14 action towards the products of ubiquitin ligases that are weak in chain-extending activity, thus adding ubiquitin to multiple sites on a target protein. Many ubiquitinated proteins can be modified at multiple lysines^{11,12,21,22}, as originally described for cell surface receptors²³. In most cases however, it is unclear whether simultaneous modification of lysines occurs on endogenous substrates, because proteomics methods for analysis of ubiquitination generally rely on proteolytic digestion. Therefore, the development of new methodologies may be critical for a better understanding of how USP14 regulates proteasome function.

We suggest that the last ubiquitin chain resists cleavage by USP14 because its site of attachment to substrate is not docked in proximity to USP14, regardless of the ubiquitin receptor to which the chain is bound. If the chain that is spared by USP14 is short, degradation will be suppressed; if it is long, then proteasome–substrate interaction will be preserved and substrate degradation will proceed. This chain will be removed by RPN11 once substrate translocation is initiated. Because RPN11 also cleaves chains proximally^{19,20}, our data reveal an unexpected point of similarity between these enzymes, which may act sequentially. However, RPN11 activity differs from that of USP14 in requiring ATP hydrolysis^{19,20}, presumably to drive translocation of the conjugate to its active site^{24,25}. That RPN11 promotes substrate degradation whereas USP14 suppresses degradation may reflect that RPN11 removes chains following the commitment step in the degradation pathway.

It will be interesting to determine whether supernumerary chain removal is the principal role of USP14 in cells, or whether other specificity rules exist for this enzyme. Substrate flexibility, a feature of the cyclin B and securin degrons²⁶, but not of ubiquitin itself, may be important for USP14 to act on these substrates. Indeed, the occlusion that we propose to impair docking of proximal ubiquitin in a free chain near the USP14 active site may apply as well to many folded globular domains that are ubiquitinated, which would be required to access the same site. This substrate docking constraint may further limit the substrate range of USP14. In contrast to USP14, RPN11 is expected to have a robust activity against chains on domains that are normally folded, because such chains may be presented to RPN11 only after the domain has been unfolded by mechanical force applied by the proteasomal ATPases.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2015; accepted 16 February 2016.

Published online 13 April 2016.

- Hanna, J. *et al.* Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* **127**, 99–111 (2006).
- Hanna, J., Meides, A., Zhang, D. & Finley, D. A ubiquitin stress response induces altered proteasome composition. *Cell* **129**, 747–759 (2007).
- Lee, B. H. *et al.* Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* **467**, 179–184 (2010).
- Peth, A., Besche, H. C. & Goldberg, A. L. Ubiquitinated proteins activate the proteasome by binding to Usp14/Ubp6, which causes 20S gate opening. *Mol. Cell* **36**, 794–804 (2009).
- Aufderheide, A. *et al.* Structural characterization of the interaction of Ubp6 with the 26S proteasome. *Proc. Natl Acad. Sci. USA* **112**, 8626–8631 (2015).
- Bashore, C. *et al.* Ubp6 deubiquitinase controls conformational dynamics and substrate degradation of the 26S proteasome. *Nature Struct. Mol. Biol.* **22**, 712–719 (2015).
- Leggett, D. S. *et al.* Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**, 495–507 (2002).
- Homma, T. *et al.* Ubiquitin-specific protease 14 modulates degradation of cellular prion protein. *Sci. Rep.* **5**, 11028 (2015).
- Hu, M. *et al.* Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J.* **24**, 3747–3756 (2005).
- Finley, D., Chen, X. & Walters, K. J. Gates, channels, and switches: elements of the proteasome machine. *Trends Biochem. Sci.* **41**, 77–93 (2016).
- Kirkpatrick, D. S. *et al.* Quantitative analysis of *in vitro* ubiquitinated cyclin B1 reveals complex chain topology. *Nature Cell Biol.* **8**, 700–710 (2006).
- Dimova, N. V. *et al.* APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1. *Nature Cell Biol.* **14**, 168–176 (2012).
- Jin, L., Williamson, A., Banerjee, S., Philipp, I. & Rape, M. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *Cell* **133**, 653–665 (2008).
- Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**, 94–102 (2000).
- Lu, Y., Lee, B. H., King, R. W., Finley, D. & Kirschner, M. W. Substrate degradation by the proteasome: a single-molecule kinetic analysis. *Science* **348**, 1250834 (2015).
- Mansour, W. *et al.* Disassembly of Lys11 and mixed linkage polyubiquitin conjugates provides insights into function of proteasomal deubiquitinases Rpn11 and Ubp6. *J. Biol. Chem.* **290**, 4688–4704 (2015).
- Vaden, J. H. *et al.* Ubiquitin-specific protease 14 regulates c-Jun N-terminal kinase signaling at the neuromuscular junction. *Mol. Neurodegener.* **10**, 3 (2015).
- Xu, D. *et al.* Phosphorylation and activation of ubiquitin-specific protease-14 by Akt regulates the ubiquitin-proteasome system. *eLife* **4**, e10510 (2015).
- Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611–615 (2002).
- Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–407 (2002).
- Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).
- Kravtsova-Ivantsiv, Y., Cohen, S. & Ciechanover, A. Modification by single ubiquitin moieties rather than polyubiquitination is sufficient for proteasomal processing of the p105 NF- κ B precursor. *Adv. Exp. Med. Biol.* **691**, 95–106 (2011).
- Haglund, K. *et al.* Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nature Cell Biol.* **5**, 461–466 (2003).
- Pathare, G. R. *et al.* Crystal structure of the proteasomal deubiquitylation module Rpn8-Rpn11. *Proc. Natl Acad. Sci. USA* **111**, 2984–2989 (2014).
- Worden, E. J., Padovani, C. & Martin, A. Structure of the Rpn11-Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nature Struct. Mol. Biol.* **21**, 220–227 (2014).
- Cox, C. J. *et al.* The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. *FEBS Lett.* **527**, 303–308 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are indebted to J. Paulo for expert assistance with mass spectrometry. We thank J. Hanna and Y. Ye for comments on the manuscript, N. Dimova for NCB1 constructs, Y. Ye for USP21 protein, Y. Saeki for PY-Sic1 clones, J. Yuan for USP14 plasmids, and R.P. Fisher and E. Winter for full-length cyclin B1 baculoviruses. This study was supported by NIH grant R01GM5660052 (to D.F.), NIH grant R37-GM043601 (to D.F.), NIH grant R01GM66492-9 (to R.W.K.), NIH grant 5R01GM039023-26 (to M. W. Kirschner), and a grant from the Rainwater Foundation to D.F.

Author Contributions B.-H.L. performed and analysed most of biochemical experiments. Y.L. performed single molecule analysis. M.A.P. performed mass spectrometric analysis. Y.S., S.S., G.T., and S.E. generated biochemical reagents. G.T. also performed structural modeling. S.P.G. guided the mass spectrometry experiments. B.-H.L., R.W.K., and D.F. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.F. (daniel_finley@hms.harvard.edu), R.W.K. (randy_king@hms.harvard.edu), or B.-H.L. (byung-hoon_lee@hms.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Recombinant proteins and biochemical reagents. Wild-type or single lysine N-terminal cyclin B1 (NCB1; N-t HA tag and C-t His tag), MBP-human E1, UbcH10, Ubc4, and Ube2S proteins were expressed in BL21 (DE3) or Rosetta2 (DE3) cells (Millipore), and purified essentially as previously described¹². The baculoviruses of full length cyclin B1, CDK1, and Cak1 (gifts from R. P. Fisher and E. Winter) were re-amplified in Sf21 cells (Life Technologies) and used to infect High Five cells (Life Technologies). Expressed proteins were purified on Ni-NTA beads (Qiagen) and by gel filtration as previously described²⁷.

N-terminal His-tagged USP14 in pET-15b (Novagen), tags-cleaved USP14 wild-type and S432E in pET-32M¹⁸ (generous gifts from J. Yuan), non-tagged USP14 and non-tagged Ubp6 in pTXB1 (NEB) were expressed in Rosetta2 (DE3), and bound and eluted from Ni-NTA beads and chitin beads (NEB), respectively, according to the manufacturers' instructions. The eluates from the resins was further applied to HiTrap Q HP5 anionic exchange chromatography (GE Healthcare) followed by gel filtration with Hi Load 16/600 Superdex 200 PG column (GE Healthcare) using AKTA Purifier FPLC system (GE Healthcare).

Sources of some commercial reagents not described elsewhere are as follows: anti-HA-HRP (3F10, Roche); anti-cyclin B1 (Ab-2, Thermo Scientific); anti-T7-HRP (Novagen); anti-Cdc27 (AF3.1, custom-made); anti-Ub (P4D1, Santa Cruz and Dako); anti-securin (MBL International); anti-Ubp6 (custom-made); anti-UbcH10 (Millipore); anti-20S (MCP21, custom-made); Ub, Ub mutants, and free Ub chains (R&D systems and Life Sensors); IU1 (custom-synthesized); o-phenanthroline (o-PA, Sigma); ATP-γS (Santa Cruz); PS-341 (Apex Bio); MG-262 (R&D Systems); USP21 (a gift from Y. Ye).

Proteasome purification and ATP-depletion. Human and yeast proteasomes were affinity-purified as previously described^{1,3}. For native human proteasomes, HEK293T cells stably expressing epitope-tagged RPN11 (a generous gift from L. Huang; mycoplasma negative) were lysed by Dounce-homogenizer in buffer A (50 mM NaH₂PO₄ (pH 7.5), 100 mM NaCl, 10% glycerol, 5 mM MgCl₂, 0.5% NP-40, 5 mM ATP and 1 mM DTT) containing protease inhibitors. The cleared lysates were incubated with NeutrAvidin resin (Thermo Scientific) for at least 2 h at 4°C. The beads were extensively washed with buffer A to remove endogenous USP14 associated with the proteasome, followed by buffer B (50 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 1 mM ATP, and 10% glycerol), then equilibrated in cleavage buffer C (50 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, and 1 mM ATP). Human proteasomes were eluted from the beads by cleavage, using TEV protease (Life Technologies) for 1 h at 30°C. For yeast proteasomes, each strain was cultured to an OD₆₀₀ of about 0.7, and lysed in buffer D (50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 5 mM MgCl₂, and 1 mM ATP) containing protease inhibitors by a French press device. The cleared lysates were incubated with rabbit IgG resin (MP Biomedicals) for 1 h at 4°C. The proteasome-bound beads were washed with an excess of wash buffer E (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 5 mM MgCl₂, and 1 mM ATP), then equilibrated in elution buffer F (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 5 mM MgCl₂, 1 mM ATP, and 10% glycerol). TEV cleavage was performed at 30°C for 1 h, and the protease was removed from the eluate using Ni-NTA resin.

Proteasomes were generally purified in the presence of ATP to preserve their integrity. To deplete the ATP from the proteasome prep where needed, purified proteasomes were incubated in reaction buffer G (50 mM Tris-HCl (pH 7.5) and 5 mM MgCl₂) containing 20 mM glucose and 0.2 U μl⁻¹ hexokinase (Sigma) for 1 h at 30°C, and then final 10% glycerol was added for low temperature storage. The hexokinase-treated 'ADP-proteasome' was tested by native gel, peptidase assay, and *in vitro* degradation assay for structural integrity and loss of activity in degradation assays.

Ubiquitin conjugate preparation and purification. Ubiquitination reactions were carried out as previously described^{3,11,12}. In summary, for APC/C-mediated reaction, APC/C was immunopurified from *X. laevis* egg extract with anti-Cdc27 antibodies bound to protein A-Sepharose 4B fast flow (Sigma). The conjugation mixture generally contained 2.5 μM MBP-E1, 3.8 μM UbcH10, 133 μM ubiquitin, 2.3 μM NCB1 or 0.55 μM full length cyclin B1-CDK1 complex, and APC/C on the beads in the reaction buffer H (20 mM Tris-HCl (pH 7.5), 100 mM KCl, 2 mM ATP, and 2.5 mM MgCl₂). Where indicated, 3.8 μM of Ubc4, 2 μM of Ube2S (along with 0.7 μM of UbcH10), or 2 μM of securin was used. For conjugation with pre-formed ubiquitin chains, 18.4 μM (K48- or K63-linked Ub₄) or 10.5 μM (K63-linked Ub₈) of free chains was incubated with 0.7–1 μM of NCB1. Reactions were typically performed at least for 90 min at room temperature with agitation at 2,000 r.p.m. The reaction mixture was collected and then combined with one bed volume of the first wash of the APC/C-beads with reaction buffer lacking salt. Sic1^{PY} ubiquitin conjugates were generated essentially as previously described³.

For TMT mass spec-coupled Ub linkage analysis and single-molecule assays, ubiquitin conjugates were further purified to remove excess unreacted free ubiquitin or ubiquitin chains and auto-ubiquitinated enzymes. The conjugation reaction was incubated with magnetic Co²⁺-beads (Life Technologies) for 30 min on ice. The conjugate-bound beads were washed two times with buffer I (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.01% Tween-20, and 1 mg ml⁻¹ BSA), then washed twice with the same buffer lacking detergent (for securin) or lacking detergent and BSA (for NCB1). Dylight⁵⁵⁰Ub_n-securin was eluted in the buffer J (50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 250 mM imidazole, and 1 mg ml⁻¹ BSA). Polyubiquitinated NCB1 was eluted first in buffer K (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 250 mM imidazole), and combined with the next two washes of the beads with buffer K lacking NaCl and imidazole to reduce the concentration of salt. The eluted conjugate was not further subjected to desalting, buffer exchange, or ultrafiltration, since all these steps resulted in a significant loss of yield and also aggregation.

Yeast strains. Yeast proteasomes with intact ubiquitin receptors were purified from strain SY409 (*MATa rpn11-TEV-ProA::HIS3 ubp6::URA3*), which derived from SDL133 (*MATa RPN11-TEV-ProA*)⁷ and SJH14 (*MATa ubp6::URA3*), the latter being prepared from a direct transformation of SUB61 with the *ubp6::URA3* cassette previously used to develop similar strains¹.

In vitro degradation and deubiquitination assays. Purified human proteasomes (3 to 5 nM) or yeast proteasomes (40 to 50 nM) were incubated with polyubiquitinated NCB1 (~110 nM) or full length cyclin B1 (~30 nM) in buffer G supplemented with 5 mM ATP at room temperature. Where indicated, purified recombinant USP14 or Ubp6 was pre-incubated with proteasome for at least 5 min before initiating the reaction. The reaction was terminated by adding 5 × SDS-PAGE sample buffer and heating for 10 min at 70°C. Samples were subsequently subjected to SDS-PAGE/immunoblot analysis. Conjugates were treated with pan-specific ubiquitin isopeptidase USP21 (0.4 μM) for 10 min. Assays with polyubiquitinated Sic1^{PY} (~200 nM) or securin (~100 nM) were performed similarly.

For USP14 deubiquitination assays uncoupled to degradation, ADP-proteasome (3 to 4 nM) was used for assays in buffer G supplemented with 3 to 5 mM 'ATP-free' ADP. To remove the residual ATP contamination, commercial ADP (Calbiochem) was first dissolved in water at 0.5 to 1 M at pH 7.0, and then treated with hexokinase and glucose in the reaction buffer as described above. The final ADP concentration in solution was adjusted to 0.25 to 0.5 M (pH 7.5). Degradation-suppressed deconjugation assays were also alternatively performed with ADP-proteasome in the assay buffer G supplemented with 3 to 5 mM ADP, 6 mM o-PA, 0.75 mM ATP-γS, 1.5 μM PS-341, and 7.5 μM MG-262, with equivalent results.

To perform deubiquitination assays with pre-formed unanchored chains, human proteasomes (5 to 12 nM) were incubated at 30°C with free chains (80 to 300 nM) in the presence or absence of recombinant USP14 (at 20-fold excess over proteasome) in buffer G supplemented with 5 mM ATP. As a control, chains were treated with USP21 (0.5 μM) for 20 min.

Tandem mass tagging mass spectrometry for Ub linkage analysis. To quantify ubiquitin conjugate cleavage by USP14, purified multi-chain NCB1 conjugates (Ub_n-NCB1 or multi-K48Ub_n-NCB1) were subjected to *in vitro* deubiquitination assays as described above. Deubiquitination reactions with ADP-proteasome in the presence or absence of USP14 were performed in duplicate and quenched by adding the inhibitor mixture of USP14 inhibitor related to IU1 (70 μM), o-PA (6 mM), ATP-γS (0.5 mM), PS-341 (1.5 μM), and MG-262 (7.5 μM). As a pan-specific deubiquitination control, USP21 (0.5 μM) was also incubated with purified conjugates for 10 min in duplicate. Each quenched reaction was subjected to overnight trypsin digestion at 37°C. Peptides obtained were desalted and labelled with TMT 10plex isobaric reagents (Thermo Scientific) following the manufacturer's protocol. Finally, all 10 samples containing TMT-labelled peptides were combined and desalted before LC-MS analysis. The peptide mixture was fractionated using a solid-phase C18 column equipped in an Accela 600 liquid chromatography (LC) pump coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific). To detect and quantify all possible linkage-specific ubiquitin peptides we designed a TMT-PRM method consisting in one full MS scan acquired in the Orbitrap (70k resolution and automatic gain control (AGC) of 1 × 10⁶) followed by 11 tMS2 (covering all detectable linkage-specific ubiquitin peptides) for each data collection cycle. Fragment ions generated by HCD (normalized collision energy was set to 27%) were acquired in the Orbitrap (70k resolution, AGC of 1 × 10⁶, 200 ms of maximum injection time and 1 m/z isolation window). RAW files obtained were processed using an in-house software pipeline based on Sequest^{28,29}. Signal to noise ratio of the reporter ions was used to calculate the abundance of each linkage across all samples.

Single-molecule deubiquitination assays. Coverslip passivation, fluorescently labelled ubiquitin-securin conjugates, TIRF microscope configuration, and image analysis were as previously described¹⁵. For single-molecule assays, 20 nM USP14-free human proteasome was incubated with 10 nM biotinylated anti-20S (MCP21)

antibody for 20 min in buffer L (50 mM Tris-HCl (pH 7.5), 5 mM MgCl₂, 3 mM ATP, 20 mM imidazole, and 3 mg ml⁻¹ BSA). After immobilizing antibody-bound proteasome on streptavidin-functionalized coverslip, unbound proteasome was washed off and securin (ubiquitinated with dylight550-labelled ubiquitin) was added at a final concentration of ~10 nM. For studying USP14-mediated deubiquitination, hexokinase-treated ADP was used, together with 6 mM o-PA. Deubiquitination kinetics was monitored under a TIRF microscope (Ti, Nikon) at 5 frames per second for 2 min.

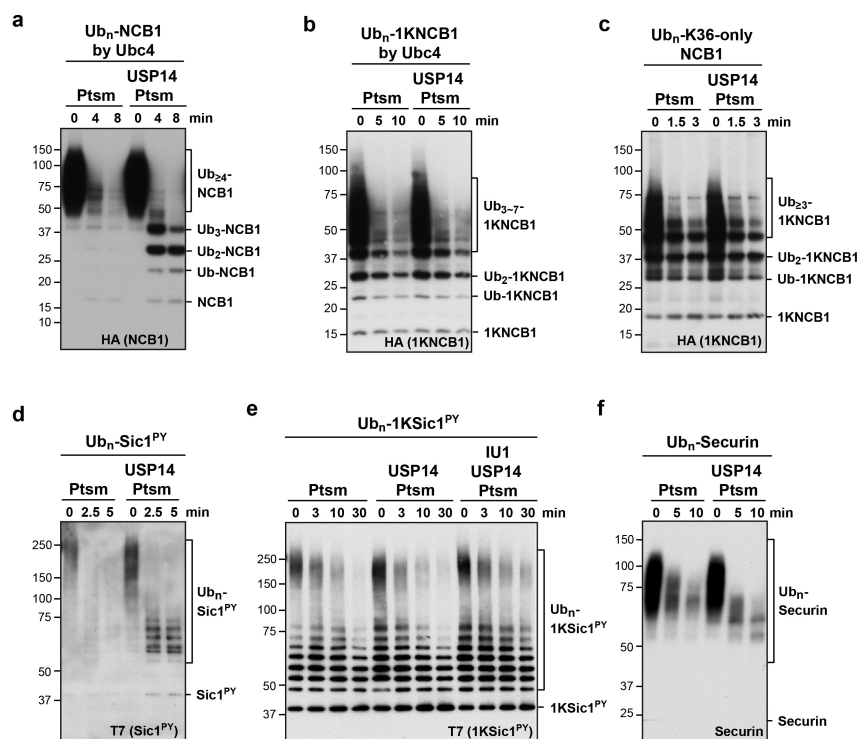
Quench-flow analysis. To resolve rapid single-encounter events for USP14-catalysed deubiquitination, millisecond reactions were performed on a quench-flow device (Kintek RQF-3). The quench-flow machine was first calibrated with a known amount of BSA, ovalbumin, and UbcH10 to achieve maximum recovery of samples according to the manufacturer's instructions. Generally, the recovery ranged between 75 to 85% of the input and was relatively even among the reaction loops (3.5 ms to minutes). For these reactions, human proteasomes were purified on a large scale as described above and concentrated to 2 µM or higher in buffer M (50 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 1 mM ATP, and 5% glycerol). Recombinant USP14 was purified and concentrated to 250 µM or higher in buffer N (35 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, and 5% glycerol). Before the quench-flow reaction, proteasomes were pre-incubated with USP14, and the solution was adjusted to 50 mM Tris-HCl (pH 7.5), 3.5 mM NaCl, 3 mM MgCl₂, 3 mM ATP, 0.74 mM DTT, and 3.3% glycerol. For the proteasome-only control (that is, no USP14), only the buffer N was added as a mock to achieve the same final composition of buffer ingredients as above. Ubiquitinated NCB1 was generated on a large scale, and before the reaction, adjusted to about 0.47 µM in buffer O (20 mM Tris-HCl (pH 7.5), 36 mM KCl, 3 mM ATP, and 3 mM MgCl₂). To enhance the recovery, prevent aggregation, and reduce non-specific binding to the tubing,

1 mg ml⁻¹ of BSA and 0.0025% of NP-40 (each final concentration after mixing with enzymes) were added to the conjugate solution. Ubiquitinated full length cyclin B1 was essentially prepared as above except adjusting the substrate at ~0.17 µM.

Millisecond single-encounter reactions were carried out from 5 ms to minutes by mixing the two reactants in the corresponding reaction loops and then quenched in the device with 3.3× quench solution P (300 mM Tris-HCl (pH 6.8), 3% SDS, and 10% glycerol) with freshly added 330 mM DTT. A small amount of SDS-PAGE loading dye was then added to each collected sample. For the 0 s sample, enzymes or ubiquitin conjugates alone were separately quenched inside the machine with each corresponding mock buffer by using reaction loop 1 at 3.5 msec. The reactions were resolved in SDS-PAGE/immunoblotting.

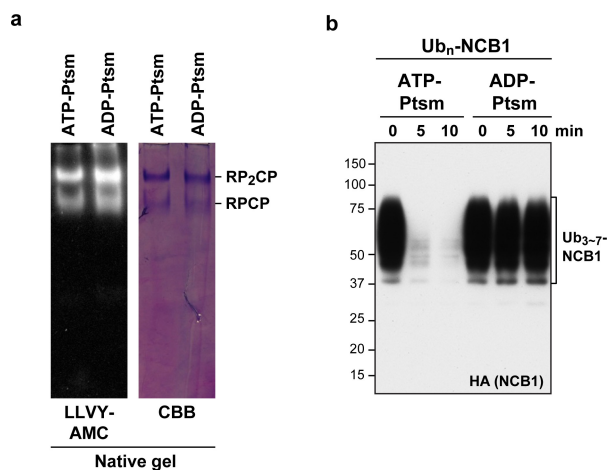
For quantification, the total ubiquitin conjugate or each conjugate species was analysed by Amersham 600 RGB system (GE Healthcare) or Odyssey CLx Infrared Imaging System (LI-COR Biosciences). The conjugate signal was normalized by UbcH10 for each time point to correct for recovery, and then further normalized to the 0 time control. For Ub_n-NCB1, proteasome alone showed some non-specific signals detected by HA antibody. This number was subtracted before normalization. The data were curve-fitted by nonlinear regression with the best fit model using GraphPad Prism (GraphPad Software).

27. Krude, T., Jackman, M., Pines, J. & Laskey, R. A. Cyclin/Cdk-dependent initiation of DNA replication in a human cell-free system. *Cell* **88**, 109–119 (1997).
28. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
29. Paulo, J. A., Gaun, A. & Gygi, S. P. Global analysis of protein expression and phosphorylation levels in nicotine-treated pancreatic stellate cells. *J. Proteome Res.* **14**, 4246–4256 (2015).

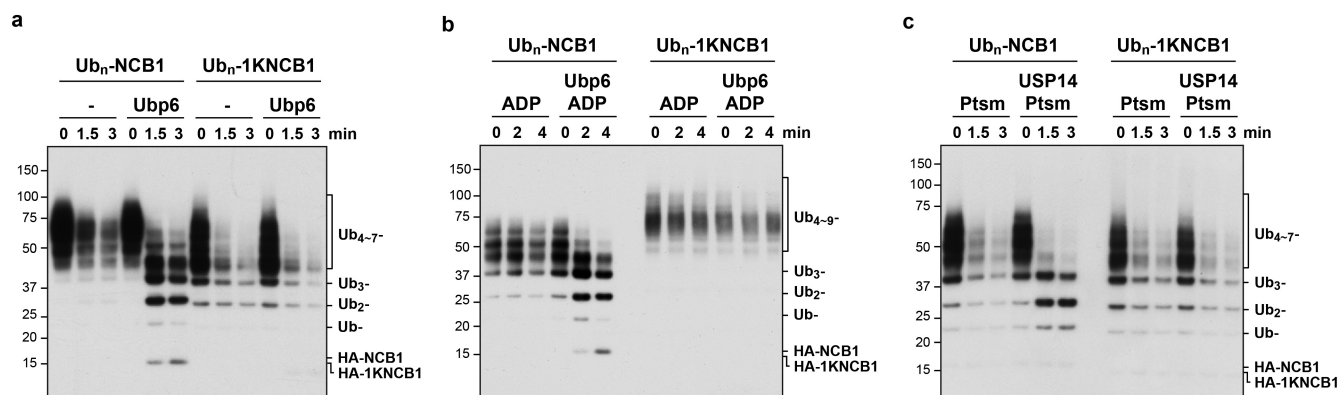


Extended Data Figure 1 | USP14 is highly specific for multi-chain ubiquitin-protein conjugates. *In vitro* degradation and deconjugation assays with polyubiquitinated conjugates. **a, b**, *In vitro* deubiquitination assays with Ub_n-NCB1 or Ub_n-1KNCB1 generated by Ubc4 in parallel. Each conjugate sample (~110 nM) was incubated with 4 nM of human proteasome in the presence or absence of USP14 (80 nM). **c**, This experiment tests whether the failure of USP14 to deubiquitinate a single-chain conjugate formed using K64-only NCB1 (Fig. 1d and f) is predictive of the behaviour of USP14 on other single-chain conjugates. Conjugates (~110 nM) were incubated with human proteasome (4 nM) in the presence or absence of USP14 (80 nM). **d, e**, *In vitro* degradation and

deubiquitination assays with polyubiquitinated Sic1^{PY} or polyubiquitinated K36-only Sic1^{PY} (1KSic1^{PY}). **d**, Ub_n-Sic1^{PY} (~200 nM) was incubated with human proteasome (5 nM) in the presence or absence of USP14 (100 nM). **e**, Assays were performed similarly as in **d**, except using a single-lysine variant of Sic1 (1KSic1^{PY}) modified with wild-type Ub. IU1 (75 μM)³ was used to inhibit USP14 deubiquitinating activity. **f**, *In vitro* degradation and deconjugation assays with polyubiquitinated securin. Human proteasome (4 nM) was incubated with Ub_n-securin (~100 nM) generated by UbcH10 and APC/C. Where indicated, USP14 (80 nM) was added. Samples were resolved by SDS-PAGE and immunoblotted using an antibody to the HA epitope (**a–c**), the T7 epitope (**d, e**), or securin (**f**).

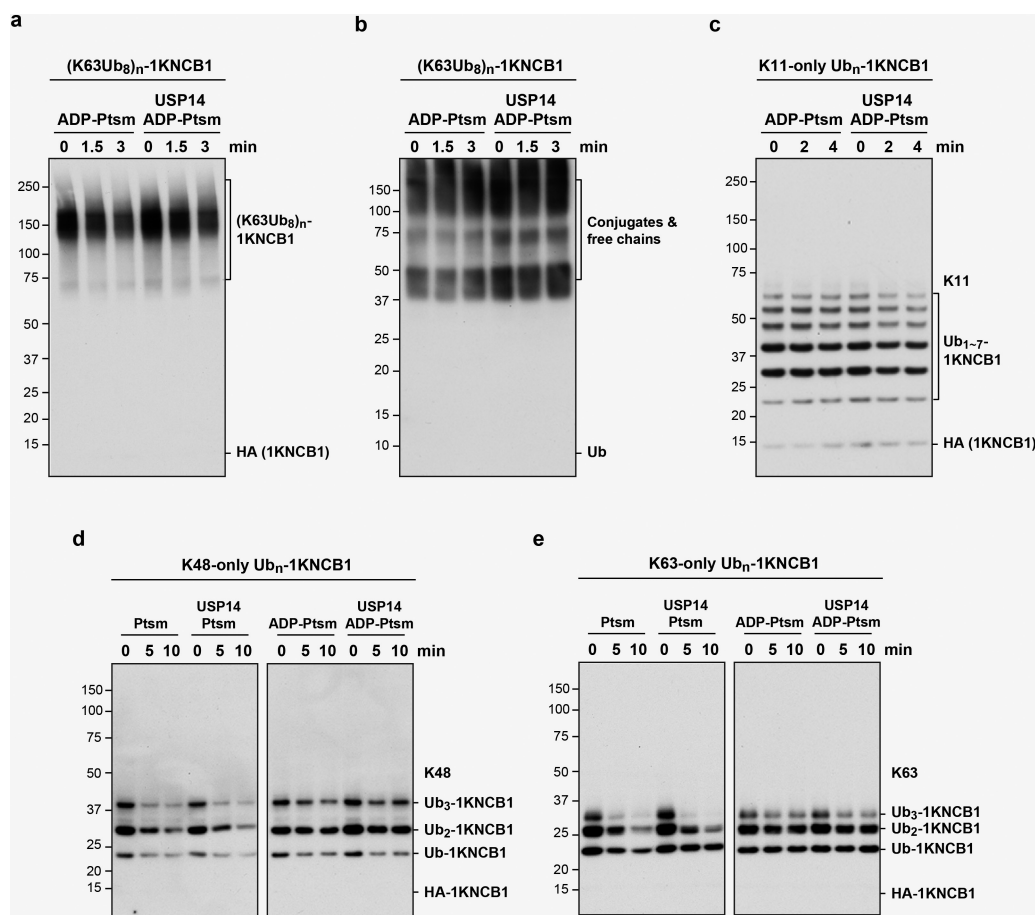


Extended Data Figure 2 | Preparation of ATP-depleted proteasome for uncoupling degradation from deubiquitination. **a**, For ATP-depletion, purified human proteasomes were treated with hexokinase or mock-treated with vehicle for hexokinase (that is, H₂O). Native gel analysis of untreated proteasome (ATP-Ptsm) or hexokinase-treated ADP-proteasome (ADP-Ptsm) with suc-LLVY-AMC and Coomassie blue (CBB) staining. Each lane was loaded with 5 µg of protein. Note that there was essentially no change in proteasome integrity upon hexokinase treatment. **b**, *In vitro* Ub_n-NCB1 degradation assays with ATP-Ptsm or ADP-Ptsm (4 nM each). Assays were performed and analysed as in Fig. 1b or e.



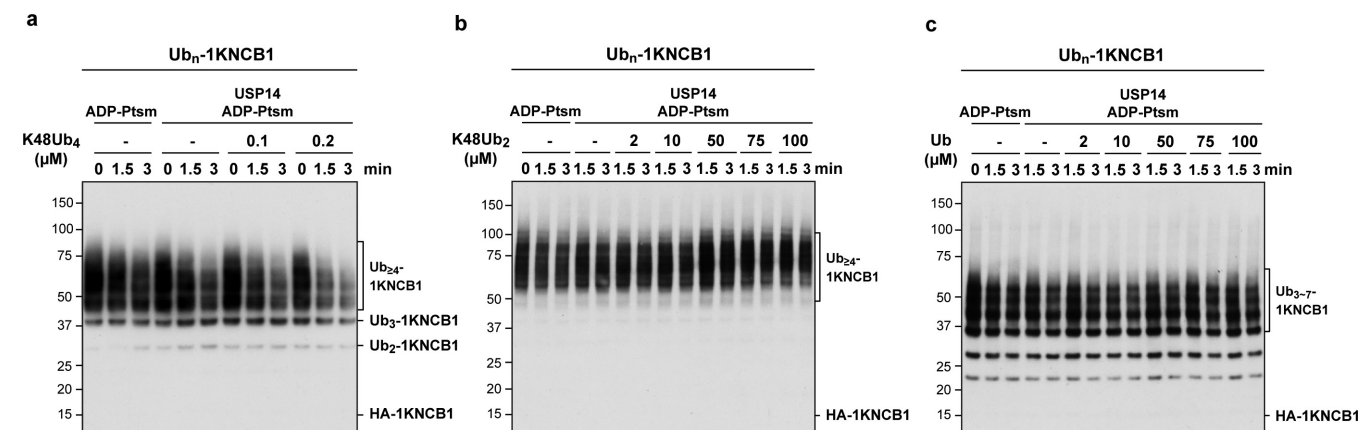
Extended Data Figure 3 | The substrate specificity of USP14 is evolutionarily conserved. *In vitro* degradation and deconjugation assays by yeast proteasome and Ubp6 with NCB1 conjugate. **a**, Purified *ubp6* Δ yeast proteasome (50 nM) was incubated with Ub_N-NCB1 or Ub_N-1KNCB1 (~110 nM) in the presence or absence of Ubp6 (200 nM). **b**, Assays as in **a**, except yeast proteasome was pre-incubated with ADP

and other inhibitors of the proteasome (ATP γ S, *o*-phenanthroline, PS-341, and MG-262) to suppress the degradation. **c**, Assays with human proteasome, USP14, and polyubiquitinated conjugates (NCB1 or 1KNCB1) were carried out as in Figs 1b and d, and compared side-by-side in one gel. These data serve as a control for Extended Data Fig. 3a, allowing Ubp6 and USP14 to be compared directly.



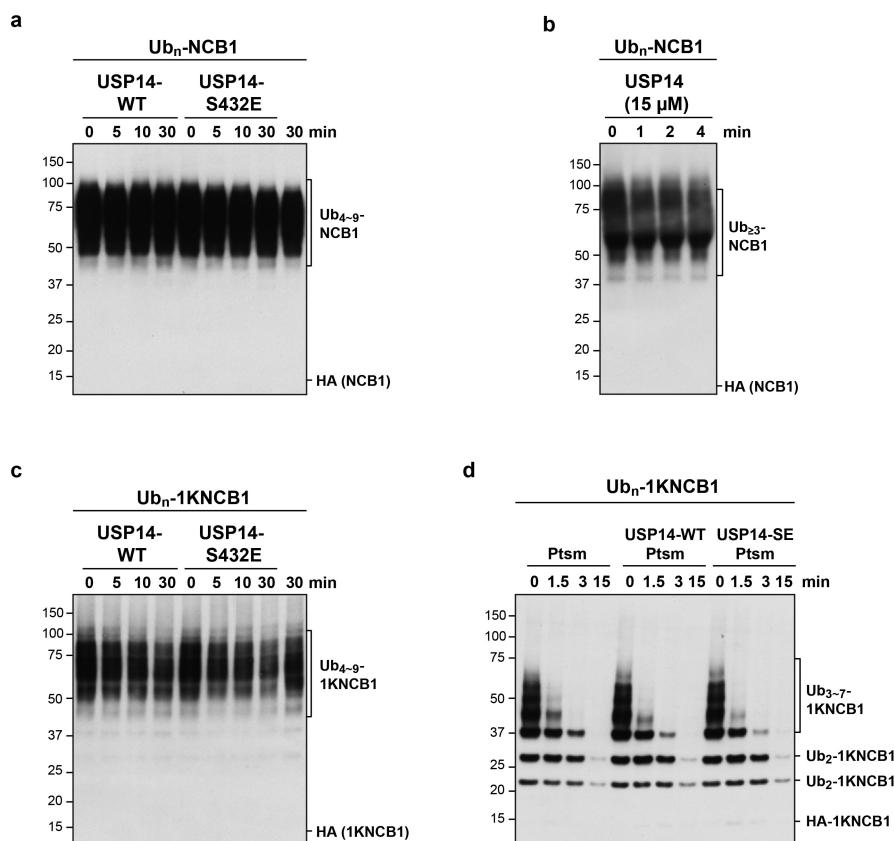
Extended Data Figure 4 | USP14 deubiquitinates singleton chains poorly, regardless of length or linkage type. *In vitro* deubiquitination assays are shown. **a**, **b**, ADP-Ptsm (4 nM) was incubated with K63-linked octa-Ub conjugates of K64-only NCB1 (~110 nM) in the presence or absence of USP14 (80 nM). **c**, Single chain conjugates were generated with Ube2S/UbcH10/APC and K11-only Ub to achieve a homogeneously K11-linked singleton species. Assays were done as in Fig. 1f. **d**, **e**, UbcH10/

APC-mediated conjugation reactions were performed on the K64-only substrate 1KNCB1 using K48-only Ub (**d**) or K63-only Ub (**e**). The conjugation reaction formed rather short chains containing no more than three ubiquitin groups. *In vitro* degradation and deubiquitination assays were performed as described above. All panels show samples resolved by SDS-PAGE and immunoblotted with HA antibody (**a**, **c-e**) or Ub antibody (**b**).



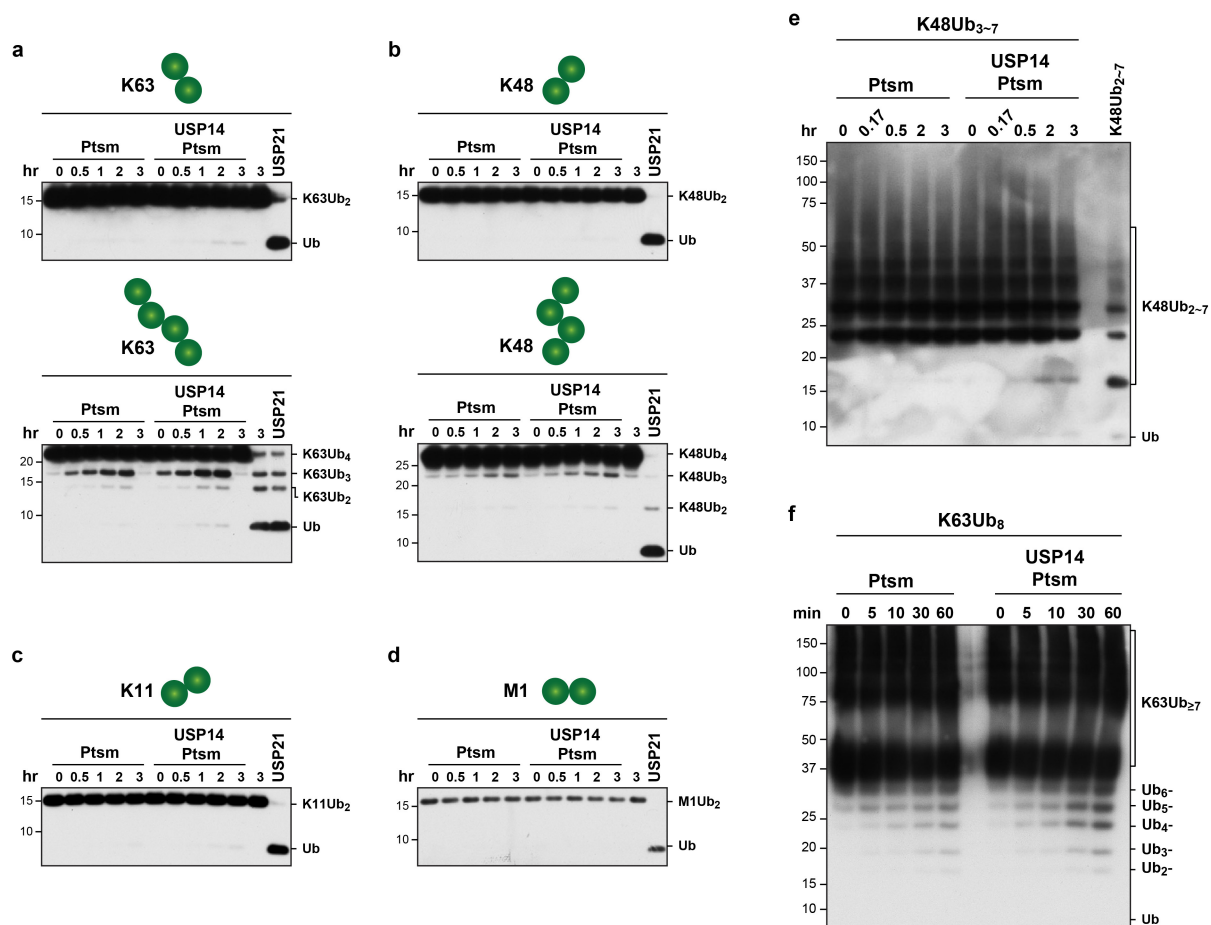
Extended Data Figure 5 | Free chains *in trans* do not stimulate USP14 activity for singleton conjugates. *In vitro* deubiquitination assays with single chain conjugates in the presence of free chains. ADP-Ptsm (4 nM) was incubated with Ub_n-1KNCB1 (~110 nM) in the presence or absence

of USP14 (80 nM), and with K48Ub₄ (a), K48Ub₂ (b), or free ubiquitin (c) at the indicated concentrations. Samples were analysed by SDS-PAGE and immunoblotted with an antibody to the HA epitope.



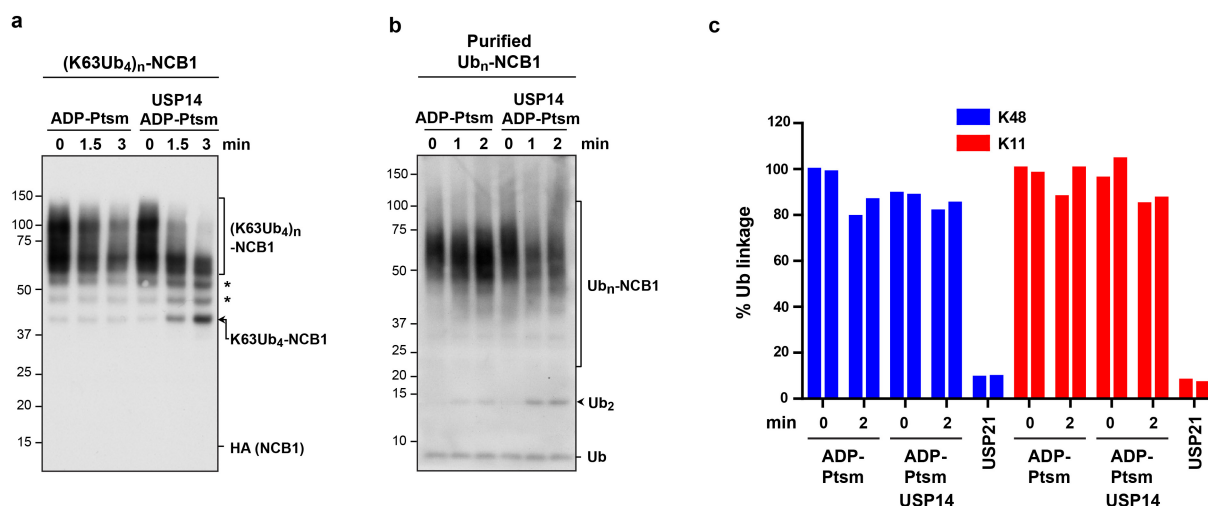
Extended Data Figure 6 | A phosphomimetic mutant of USP14 that mimics AKT-phosphorylated enzyme does not promote cleavage of singleton ubiquitin chains. **a**, Free USP14 wild-type (400 nM) does not cleave multi-chain cyclin B1 conjugates (~110 nM). **b**, High concentration of free USP14 does not deubiquitinate Ub_n-NCB1. USP14 (15 μM) activity on Ub_n-NCB1 (~0.45 μM) in the absence of proteasome was tested by incubating at room temperature for time periods as indicated. Note that there is no obvious deubiquitination of the conjugate by free USP14, although high levels of USP14 were tested (>1,000-fold over

that of activated, proteasome-associated USP14 in our standard assay). This experiment confirms that deubiquitination of Ub_n-NCB1 by USP14 occurs on and requires the proteasome. **c**, Free USP14(S432E) (400 nM) does not cleave single chain cyclin B1 conjugates (~110 nM). **d**, Proteasome-associated USP14(S432E) (4 nM proteasome and 80 nM USP14) does not cleave single chain cyclin B1 conjugates (~110 nM). Assay in **d** was performed in the presence of ATP under conditions permissive for degradation. Samples were resolved by SDS-PAGE and immunoblotted using an antibody to the HA epitope.



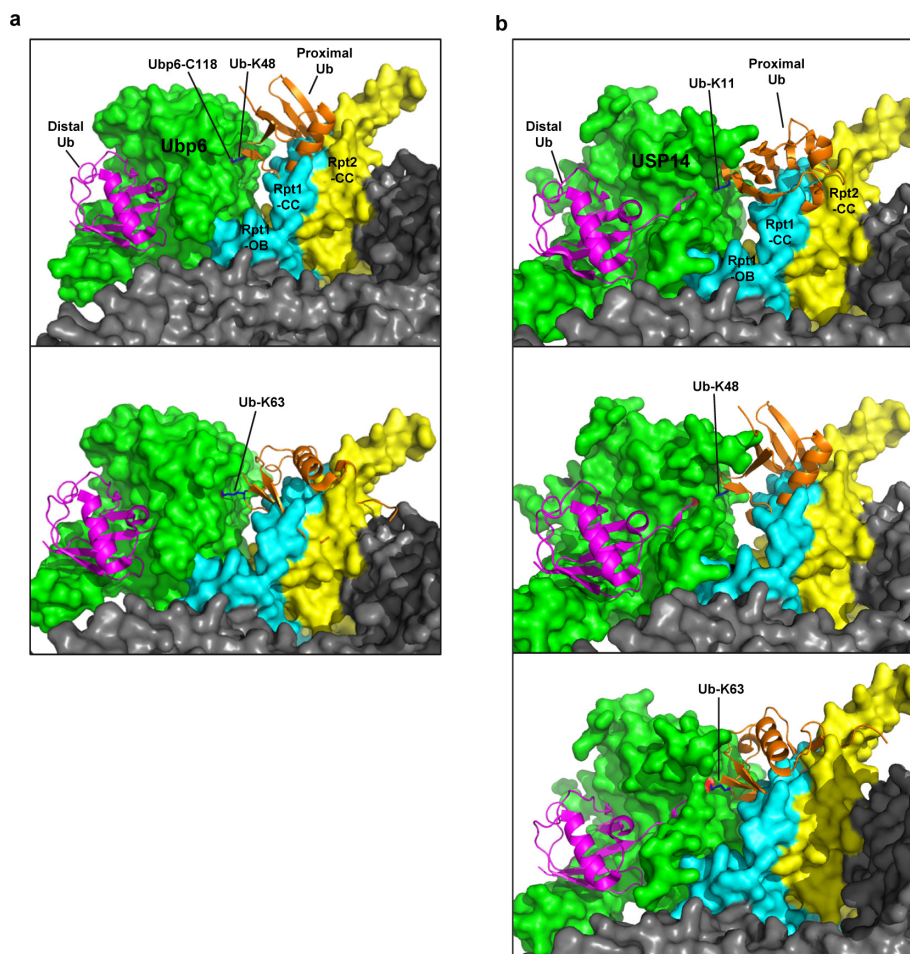
Extended Data Figure 7 | USP14 deubiquitinates free chains poorly, regardless of linkage type. *In vitro* deubiquitination assays with free chains. **a–d**, Each linkage-specific free chain (200 nM) was incubated with human proteasome (12 nM) in the presence or absence of USP14 (240 nM). The sample was treated with pan-specific Ub isopeptidase USP21 (400 nM) for 20 min as a positive control showing that chain

disassembly is readily detected. **e, f**, K48-linked Ub₃₋₇ free chains (125 nM) or K63-linked octa-Ub free chains (300 nM) were incubated with human proteasomes (5 nM for **e** and 10 nM for **f**) in the presence or absence of USP14 (20-fold excess over proteasome). Samples were resolved by SDS-PAGE and immunoblotted using an antibody to ubiquitin.



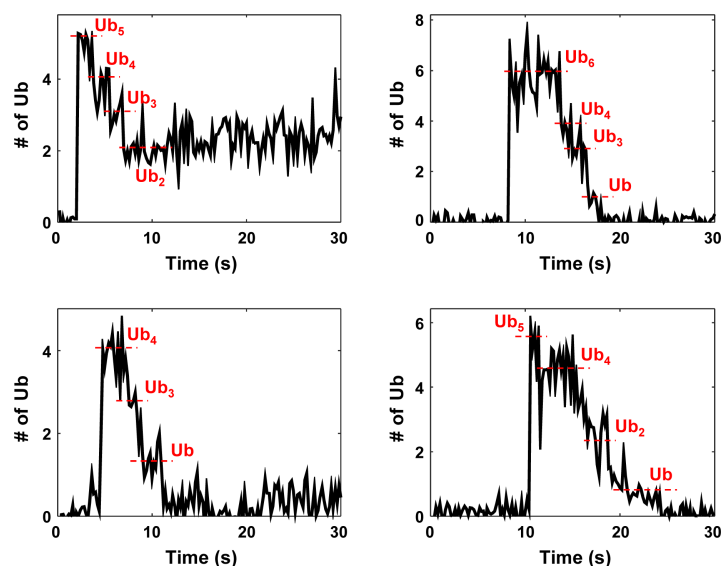
Extended Data Figure 8 | USP14 cleaves ubiquitin chains at the proximal site. *In vitro* deubiquitination assays with multi-chain conjugates. **a**, Multi-chain K63Ub₄ conjugates on NCB1 were prepared using pre-formed K63Ub₄ free chains as described in Methods. Assays were performed as in Fig. 2a. Arrow indicates accumulated K63Ub₄-NCB1 as major deubiquitinated species generated by USP14 cleavage. Asterisks indicate presumptive deconjugated K63Ub₅- or other multi-chain K63Ub_n-NCB1 species originating from other K63Ub_n contaminants in the commercial preparation of K63Ub₄. **b**, Ub_n-NCB1 was generated using UbCH10, APC/C, and wild type ubiquitin, and purified as described in Methods. Note that this conjugate may be enriched with short

multi-ubiquitin chains with different linkages, based on previous reports¹². Deubiquitination assays were performed as in Fig. 2c. Arrowhead indicates di-Ub chains as major products of USP14 activity, suggesting that Ub_n-NCB1 in this sample was rich in diubiquitin chains. Samples were resolved by SDS-PAGE and immunoblotted with antibody against the HA epitope (**a**) or Ub (**b**). **c**, TMT labelling and quantitative mass spectrometry. Purified Ub_n-NCB1 was assayed in biological duplicates for each condition, and the resulting samples were analysed for specific ubiquitin chain linkages by mass spectrometry as in Fig. 2d. The experiment was repeated in biological duplicates at a 4-min point with equivalent results (data not shown).



Extended Data Figure 9 | Modelling of free ubiquitin chain cleavage by proteasome-activated Ubp6 or USP14. Possible steric conflict between the proximal ubiquitin and proteasome-associated Ubp6 or USP14 was assessed by molecular modelling. **a**, As in Fig. 2f, except employing K48- or K63-linked proximal ubiquitin. **b**, Proximal ubiquitin with each different linkage was fitted in USP14 structure with minimal steric

hindrance, and then docked onto the proteasome electron microscopy structure. Structures were obtained from the PDB database: Ubp6 and proteasome (5A5B), USP14 (2AYO), K63 di-Ub (2RR9), and ubiquitin (1UBQ). Note that for K6, 27, 29, and 33 linkages, modelling on the proteasome was not attempted because of extensive steric occlusion between proximal ubiquitin and USP14.



Extended Data Figure 10 | Single-molecule studies of RPN11 show patterns of chain removal similar to those of USP14. These experiments serve as controls for Fig. 3. Since RPN11 is known to work as a proximal-ubiquitin-specific DUB, its similarity to USP14 in these particular chain removal assays supports the hypothesis that USP14 is a proximal-specific deubiquitinating enzyme. Single-molecule assays were performed and analysed as described in Fig. 3, and were carried out using the same

samples of proteasomes and ubiquitinated securin. The data shown are examples of step-trace segmentation graphs from a single-molecule analysis of purified Dylight550-labelled Ub_n -securin conjugates incubated with human proteasome lacking USP14 in the presence of ATP. These deubiquitination events represent RPN11-mediated activity as previously reported¹⁵.

CORRIGENDUM

doi:10.1038/nature16537

Corrigendum: D14–SCF^{D3}–dependent degradation of D53 regulates strigolactone signalling

Feng Zhou, Qibing Lin, Lihong Zhu, Yulong Ren, Kunneng Zhou, Nitzan Shabek, Fuqing Wu, Haibin Mao, Wei Dong, Lu Gan, Weiwei Ma, He Gao, Jun Chen, Chao Yang, Dan Wang, Junjie Tan, Xin Zhang, Xiuping Guo, Jiulin Wang, Ling Jiang, Xi Liu, Weiqi Chen, Jinfang Chu, Cunyu Yan, Kotomi Ueno, Shinsaku Ito, Tadao Asami, Zhijun Cheng, Jie Wang, Cailin Lei, Huqu Zhai, Chuanyin Wu, Haiyang Wang, Ning Zheng & Jianmin Wan

Nature **504**, 406–410 (2013); doi: 10.1038/nature12878

In this Article, the bottom panel of Fig. 3c was assembled from processed digital images of separate western blot autoradiographs to show approximately equal amounts of MBP–D3 protein input in the presence or absence of GR24 treatment. During figure preparation, a single MBP–D3 lane was duplicated in error to create the third and fourth lanes. In addition, the last set of images in Extended Data Fig. 8b, right panel, was used in error for Extended Data Fig. 8c, bottom left panel. We have now repeated the relevant experiments, and Fig. 1 of this Corrigendum shows the amended Fig. 3c, and the Supplementary Information to this Corrigendum shows the corrected Extended Data Fig. 8c. None of these corrections alters the original design of the experiments, results, interpretation or conclusions of the paper. We apologize for any confusion that this may have caused.

Supplementary Information is available in the online version of the Corrigendum.

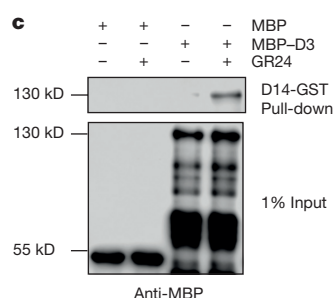


Figure 1 | This is the corrected Fig. 3c of the original Article.

CORRIGENDUM

doi:10.1038/nature16499

Corrigendum: CMT2D neuropathy is linked to the neomorphic binding activity of glycyl-tRNA synthetase

Weiwei He, Ge Bai, Huihao Zhou, Na Wei, Nicholas M. White, Janelle Lauer, Huaqing Liu, Yi Shi, Calin Dan Dumitru, Karen Lettieri, Veronica Shubayev, Alben Jordanova, Velina Guergueltcheva, Patrick R. Griffin, Robert W. Burgess, Samuel L. Pfaff & Xiang-Lei Yang

Nature **526**, 710–714 (2015); doi:10.1038/nature15510

The Acknowledgements of this Letter should have included the following sentence: “This work was also supported in part by the University of Antwerp (TOP BOF 29069 to A.J.); the Fund for Scientific Research-Flanders (grant number G078414N to A.J.); the Association Belge contre les Maladies Neuromusculaire (to A.J.); the Association Française contre les Myopathies (grant number 16197 to A.J.).

ERRATUM

doi:10.1038/nature16541

Erratum: Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation

Ingrid E. Wertz, Kim Newton, Dhaya Seshasayee, Saritha Kusam, Cynthia Lam, Juan Zhang, Nataliya Popovych, Elizabeth Helgason, Allyn Schoeffler, Surinder Jeet, Nandhini Ramamoorthi, Lorna Kategaya, Robert J. Newman, Keisuke Horikawa, Debra Dugger, Wendy Sandoval, Susmith Mukund, Anuradha Zindal, Flavius Martin, Clifford Quan, Jeffrey Tom, Wayne J. Fairbrother, Michael Townsend, Søren Warming, Jason DeVoss, Jinfeng Liu, Erin Dueber, Patrick Caplazi, Wyne P. Lee, Christopher C. Goodnow, Mercedesz Balazs, Kebin Yu, Ganesh Kolumam & Vishva M. Dixit

Nature **528**, 370–375 (2015); doi: 10.1038/nature16165

In this Article, owing to a typesetter error the ‘received date’ was incorrectly shown as ‘5 November 2015’ instead of ‘5 November 2013’; this has been corrected in the online versions of the paper.

CAREERS

MISCONDUCT Advice for would-be whistle-blowers **p.405**

DIVERSITY Scientists combine research with family life go.nature.com/cdfpaj

NATUREJOBS For the latest career listings and advice www.naturejobs.com



The silver-spotted skipper (*Hesperia comma*) is one of many species adapting to changing conditions.

ECOLOGY

Change is in the air

Animals and plants worldwide are responding to shifts in the environment — and so must the scientists who study them.

BY BARBRA RODRIGUEZ

Reptile researcher Russell Burke saw 14 years of fieldwork nearly wiped out in a moment, when Hurricane Sandy hit the US east coast in 2012. Since 1998, he had been studying the population dynamics of two native species of turtle — their clutch size, seasonal movements and more — at Jamaica Bay Wildlife Refuge in New York City, near his workplace at Hofstra University in Hempstead. But in 2012, storm surges breached the

wall of a pond that was home to 22 snapping turtles (*Chelydra serpentina*) in his study, and all were killed when the waves turned the pond from freshwater to salt. The region's diamondback terrapins (*Malaclemys terrapin*), which prefer brackish waters, fared better there. Burke noted intriguing responses. "They started nesting in places they hadn't before, and stopped in others where they had," he says. But the breach lost him easy access to the terrapins' nesting area, so his team could examine only a fraction of the nests.

It will be years yet before the wall is repaired, salinity drops and the snapping turtles are reintroduced. It will take longer still for Burke and his team to assess the storm's full impact on both species' populations. It is a blow to his turtle research programme. "We're going to have this big gap," he says. "Until they fix the breach, we can't figure out what's really going on."

Shifting weather patterns — milder winters, wetter springs and storms that are more frequent and more severe — are increasingly changing the landscape for scientists who study flora and fauna in the field (see 'In the face of unpredictability'). Glacier lilies in the US mid-west are blooming before their hummingbird pollinators arrive¹; malnourished Atlantic salmon are entering Scottish waterways later²; birds and plants are moving higher in the Swiss Alps³; and amphibians in drought-stricken California are struggling to stay alive⁴. All over, environmental changes are forcing plants and animals to modify their survival tactics.

Ecologists, wildlife biologists, behaviouralists and others who do fieldwork must in turn respond and be flexible. They might have to change how they evaluate habitats during field surveys, head into the field earlier or factor more staff and equipment into their budget. If they visit remote regions, they may need to pay more attention to their physical safety as extreme weather events become more violent or frequent, and as changing conditions bring greater threat of wildlife encounters.

BEAR SCARE

For some, those encounters can be dangerous indeed. George Divoky, a climate-change ecologist at the University of Alaska Fairbanks, routinely considers personal safety now as he studies black guillemot (*Cepphus grylle*), a type of Arctic sea bird that breeds on a barrier island off Alaska's north coast. In 2002, a search-and-rescue squad had to collect him and his team in a helicopter because polar bears moved in to their camp as a result of ice melt. In earlier decades, only solitary bears had been spotted in the distance on the 5-kilometre-long island. That year, more than two dozen went rummaging through the scientists' tents over a four-night period. "I carry a loaded shotgun all the time now when I go to the island," says Divoky.

As the geographic ranges of studied species shift or become more varied or diffuse, researchers might need more or larger field sites, which requires more travel and hands-on help. Robert Curry, a species-hybridization specialist at Villanova University in ►

► Pennsylvania, found this out in 2003. He works on the causes and consequences of interbreeding between two chickadee species with overlapping habitats. When he started his research in 1998, he had a field site in a spot where the ranges intersected, as well as single-species zones to the north and south. But by 2003, he and his team had begun to notice changes in the most northerly area, including the emergence of birdsong associated with the southern chickadee species, which had shifted its range northward to adapt to warmer winters⁵.

As a result of that migration, Curry had to add a fourth study site 24 kilometres above the most northerly area to single out a pure population. Each spring, he travels from Villanova, 48 kilometres below the southernmost site, to 200 nests across his 4 field sites. If he had time to visit all of them in one trip, he would end up travelling more than 160 kilometres to study adults and nestlings. Curry estimates that the extra site has added 160 hours of fieldwork annually.

Ecologist Chris Thomas at the University of York, UK, had to add staff members after one of the butterfly species that he investigates shifted its location. The silver-spotted skipper (*Hesperia comma*) — once restricted to south-facing hillsides — had moved by the end of the 1990s to taller, shadier vegetation on hillsides that face east and west, and now even on some that face north⁶. “There is a lot more land to cover,” he says.

When Thomas launched his programme in 1982, he needed only one lab member to survey the butterfly’s habitat. But by 2009, that number had ballooned to 12. And he needs more help yet: he now relies on field-site temperature gauges and on an ecologist colleague to model habitat temperatures in the butterfly’s range.

Although his fieldwork and analysis have grown more complex, Thomas says that it’s

Researchers who work in the field need to appreciate — and be ready for — the effects of changing environments.

- Understand the risk of experiments going awry, if your study subject is vulnerable to climate events. That might be true for population sizes that are small, or if a species has experienced excessive habitat fragmentation or has low genetic variability.
- If possible, research a site that already has long-term climate and ecological data, or that covers a large area. Having a larger site or more sites and time for research improves the chance of gaining meaningful data about the biological trait under study, despite weather anomalies. This also makes it easier to see whether climate events affect your study subject, although it might require more staff and travel.

SCIENTIFICALLY FIT

In the face of unpredictability

- A range of tips and case studies on the impact of extreme climate events on ecological studies are available in the *Journal of Animal Ecology* in a paper⁷ by climate-change biologist Liam Bailey and climate-change ecologist Martijn van de Pol, both at Australian National University in Canberra.
- To better understand the broader impacts of climate change, review reports from the Intergovernmental Panel on Climate Change (IPCC). The IPCC recently published a report⁸, for instance, about extreme climate events and their mitigation that discusses temperature and precipitation predictions across different regions.
- For more-targeted information, consult national weather data sets and primary literature about historic and predicted changes in your area. **B.R.**

important to have longitudinal data. “The real value is keeping the research going and understanding those long-term changes,” he says.

THWARTED PLANS

Some scientists have to shift focus entirely when conditions change. In 1998, Paul Dolman, an applied ecologist at the University of East Anglia in Norwich, UK, began to explore how and why the habitat of woodlarks was changing in the 19,000-hectare Thetford Forest, a plantation in Breckland, UK. The UK Forestry Commission has managed and logged the forest for decades to support Corsican pine (*Pinus nigra* ssp. *lario*), which thrived in the region’s warm, dry climate. In 2000, the commission began to discuss designating the forest as a conservation site for woodlark (*Lullula arborea*), whose numbers had peaked at 456, and the plan was finalized in 2006. Despite special protection, the woodlark population has since declined to less than one-third of that number. The question is: why?

Woodlark, surveys found, prefer short grasses growing in spots where Corsican pine has been felled. But Dolman, who had pored over a century’s worth of precipitation and 60 years of temperature records for the district, saw that the climate was becoming milder and wetter. His preliminary findings showed that vegetation in the clearings was growing more quickly, which rendered the areas uninhabitable to woodlark. He set out to clarify whether climate-driven vegetation changes were involved in the decline.

He never got an answer — in 2008, he was forced to change gears. The forest’s increasingly wet conditions nurtured the growth of a fungus that began to stunt Corsican pine growth, and the commission began to replace the trees with hemlock, fir and other species. Dolman and his team have since begun to study the

woodlarks’ habitat preferences among the new trees, and how the forest should be managed to keep the bird population healthy. “Be ready to think on your feet,” he warns.

The cost of environmental changes can also add up. Divoky has had to raise more than US\$32,000 to buy extra equipment and supplies as a direct result of the changing environment. Among his purchases are 200 plastic cases for the guillemot, after bears learned in 2009 how to open wooden nest boxes and ate all but one fledgling; an electric fence to deter the bears; a freezer, because the weakened permafrost can no longer support a freezer pit; and rain gear to endure more-violent and frequent rainstorms.

He does not know what he will face this spring during his annual trek to the island. “But,” he says, “I’m prepared to deal with it.” ■

Barbra A. Rodriguez is a freelance writer in Austin, Texas.

1. McKinney, A. M. et al. *Ecology* **93**, 1987–1993 (2012).
2. Windsor, M. L., Hutchinson, P., Hansen, L. P. & Reddin, D. G. *Atlantic Salmon at Sea: Findings from Recent Research and their Implications for Management* (North Atlantic Salmon Conservation Organization, 2012).
3. Roth, T., Plattner, M. & Amrhein, V. *PLoS One* <http://dx.doi.org/10.1371/journal.pone.0082490> (2014).
4. California Department of Fish and Wildlife. *A Rapid Assessment of the Vulnerability of Sensitive Wildlife to Extreme Drought* (CDFW, 2015).
5. Taylor, S. A. et al. *Curr. Biol.* **24**, 671–676 (2014).
6. Lawson, C. R. et al. *The Status and Conservation of the Silver-Spotted Skipper Hesperia comma in South-East England 2000–2009* (Univ. Exeter, 2013).
7. Bailey, L. D. & van de Pol, M. J. *Animal Ecol.* **85**, 85–96 (2016).
8. IPCC. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2012).



Ecologist Paul Dolman in the field in Uzbekistan.

COLUMN

How to survive as a whistle-blower

Scientists who report potential misconduct must tread carefully, says **Michael Doran**.

In 2012, a graduate student came to me for advice. His supervisor was travelling, and the student needed images for a review article that he was writing. I suggested that he check his lab's shared folders for something suitable. When he came to me again, he was shaking.

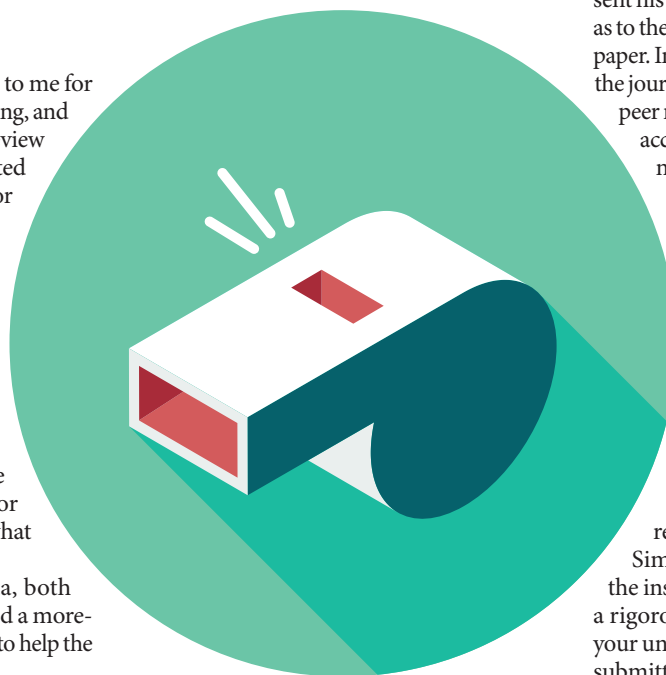
The student had found images and data that he was very familiar with, but that bore unfamiliar descriptions. He was concerned that findings that he had been unable to replicate from a previous publication, and which were the foundation of his PhD project, were misleading. He wanted me, a junior member of the faculty, to tell him what to do next.

I spent hours looking at the data, both with him and on my own. I consulted a more-experienced colleague, and I decided to help the student to submit a complaint.

Ultimately, I advised that he take his concerns to the vice-chancellor. Although we did not know it at the time, doing so mandated that the complaint be formally investigated. It also gave the student some protection from the possible repercussions of being a whistle-blower.

Internal investigations followed, as well as a retraction, repayment of a major national grant and an external inquiry that found research misconduct and multiple misrepresentations. The vice-chancellor violated state policy by disclosing my name to a reporter. Although he self-reported the breach, I wondered what else was being said about me and to whom. The difficult situation was new to us all. It affected my health, productivity and relationships; I lost countless hours that I could have devoted to lab work. The student, who has still not graduated, is taking a break from research.

There is no handbook that describes what to do in these situations. If you decide to be a whistle-blower, you must realize that it will be stressful. And because it is so stressful, you want to ensure that any investigations that are carried out will be robust. Every case needs to be considered on an individual basis, but I hope that sharing my recommendations will help others who find themselves in a similar position.



Don't confront potential misconduct alone.

Although postdocs and PhD students are the most likely to identify inconsistencies in the previously published data of their groups, they are often the least equipped to highlight serious problems. Be ready to give your supervisor the benefit of the doubt, but also be aware that raising concerns directly could provide an opportunity to obscure evidence of misconduct. In fact, the co-founders of the blog Retraction Watch recommend against contacting authors first if no one else knows of issues in the research.

Before submitting a complaint, ask a technical expert (preferably from outside your university) to corroborate your assessment. In my case, I asked others to verify my evaluation of data without disclosing exactly why. You must be absolutely confident of the veracity of your complaint and be ready for your expertise to be called into question.

Make your case clearly and keep detailed records. Describe inconsistencies thoroughly, and ensure that any correspondence is recorded electronically; complaints with a digital trail are less likely to disappear. Maintain careful records, and retain all of the original data, if possible.

Do not relinquish these records, and neither deny nor admit that you have them. Doing so could mean that authorities might either dismiss your assertions or accuse you of holding unauthorized copies of data.

Submit your complaint to the highest authority. A stifled complaint will degrade your credibility. Take steps to maximize the chance that your concerns will be investigated and that you will be protected from retribution. In our case, an initial internal inquiry did not result in corrective actions. Frustrated, the student then sent his complaint to Retraction Watch, as well as to the journal that had published the original paper. In the retraction statement that followed, the journal editor said that he did not think that peer reviewers would have deemed the paper acceptable had they known the "extent and nature of the mistakes" that it contained. A similar scenario occurred with the grant.

In retrospect, I would recommend that you submit complaints directly to journals and funding agencies. Although journals cannot conduct investigations themselves, they can demand further information and data from authors. They are also best equipped to evaluate the significance of errors within their publications and have the power to retract papers when flaws are revealed. Similarly, funding agencies can instruct the institutions that they support to conduct a rigorous review. Even though members of your university are likely to learn or guess who submitted the complaint, the involvement of external organizations creates a buffer between you and the institution.

Avoid public disclosure. In my view, it is not appropriate to make public statements about such cases until they are resolved. It would have been much easier for me if our case had not been discussed openly until its outcome had been decided. However, I think the student found that talking to the press helped to gain traction within the university.

Despite the stress it caused and the time it required, I would still urge other scientists to help bring misconduct to light if they uncover it. As more cases emerge, it becomes easier for other whistle-blowers to come forward: this incident prompted the university to implement progressive policies concerning responsible research practices. Supporting scientists and protecting scientific integrity are part of our obligations as academics. Misconduct hurts science and wastes taxpayers' money. Remember — you are doing the right thing. ■

Michael Doran is a stem-cell scientist at Queensland University of Technology in Brisbane, Australia.

ILLUSTRATION BY J. VAST

THE PARASITE AND THE WIDOW

A way of life.

BY JEREMY M. GOTTWIG

The taxi driver tossed Sheli's bags through the gate. "I've got a long flight back," he grumbled. "A long flight." He closed the hatch. The taxi nudged itself away from the small habitat. It arced back towards the peridot planet of Caldera: Sheli's home for 61 years.

She felt lonely. It hurt to leave.

The taxi driver hadn't permitted a lick of conversation that didn't involve his own complaints. Sheli had wanted to talk about her life and experiences as the only Earth human among a collective of worion expats. She wanted to ramble on about her husband, her other lovers, and her husband's other lovers. She wanted to ruminate about their increased solitude as the collective peeled away. She wanted to talk about her husband: his awkward art and fondness for annoying pets.

She wanted to talk about everything but his death.

Sheli sat on the narrow bench, stared out into silent space and waited for her shuttle.

She heard something: a mechanical whisper followed by three clicks. Sheli noticed an elegant tripod at the opposite end of the small shuttle stop. The tripod carried a glass cylinder of pink gel. A black, serpentine creature with a mouth at each extremity hung suspended in that gel.

"I startled you," came a monotone, synthetic voice.

"You did," Sheli confirmed. She tried to ignore her companion, but Sheli was never one to keep silent company. "Are you travelling to Earth as well?" she asked.

"Sadly, no." The tripod took a few steps towards her. "I am returning home."

"And where is that?"

A pause and then: "THREE."

Sheli gasped. She glanced at the hatch and searched space for signs of her shuttle but saw only stars. With a nervous sigh, she said: "So I'm trapped in a bubble with a predatory parasite. You are a clever hunter."

"You have nothing to worry about from me."

Sheli frowned. She fought the urge to scoot away from the tripod. "I suppose I'm too old for you?"

"Not at all, but we would both be required to fill out the proper paperwork. Then come integration procedures, and I suspect we would need to go through a batch of menopause-reversal treatments. Only then can

I attach." The parasite paused and added: "We are desperate for hosts, but I'll admit to some irony that we have yet to streamline this whole process."

Sheli winced and attempted to change the subject. "How long have you been away from THREE?"

"Not long. I had a line on a potential host, but it didn't work out."



"But Caldera seems so far away."

"I am nearing the end and growing desperate. I received a relay about a boy with a terminal illness. My enzymes would have renewed his body, but his parents had second thoughts and sent me away. I understood their concern, but I wasted a lot of time filling out the forms required just to enter Xevilius. He would have lived for more than a thousand years and had a full life. By now, he is probably dead."

"Well, why not cure the boy and then detach yourself?"

"I cannot detach myself until I have laid and fertilized my eggs. If I tried, both the host and I would die. When you join with the Parasite-S, you must remain with the S."

"But you kill the host eventually? No matter what?"

"I must eventually reproduce. The host becomes a vessel for my offspring. I admit that the whole process must seem disgusting."

"Most sex is, I suppose. How much time do you have left?"

"Not long."

Sheli felt an uncomfortable kinship with the creature.

They were both heading home just as age prepared to get the better of them.

"Earth is your home?" the parasite asked.

"Yes. I'll be living with my sisters. Once I have apologized for a few things, that is." She paused and traced a series of stars with her eyes. "I haven't been home in over 60 years." A thought occurred to her. She chuckled. "Can you make me young again?"

"Between my enzymes and our treatments, you would be renewed." The parasite paused, stepped closer, and added: "I promise you a long, experiential life."

Sheli sighed. She turned back to the stars and let her words wander. "I came to Caldera, because I wanted experiences. That was — oh my — 20 years after First Contact. It seemed like not much had changed on Earth. Grow up. Go to school. Get married. Have kids. Get old. Die. Out there were new worlds; unusual people; fascinating sexual ethics; surprising visions of beauty."

She began flipping through pages in her notebook and found a sketch of her worion husband.

She held up the notebook. "Can you see?"

"Heat signatures only."

"Oh..." She closed the notebook and added: "You say that you can make me young. What I really need is for someone to bring me back. To the beginning. So that I can live it all again."

"There are new loves to be had."

Sheli considered this. She shook her head. "Not for me, I think. Not anymore."

"I understand. You can't fault me for trying."

Sheli smiled at the parasite and said nothing.

A few moments later, she noticed a flash out of the corner of her eye. A shuttle came alongside the bubble and docked.

"I believe this is mine," the parasite said as the hatch whirled open.

"Best of luck," Sheli said. The words embarrassed her.

"I wish you and your sisters the best," the parasite said before climbing on board. Moments later, the shuttle detached and drifted away from the bubble. Sheli watched it heel to port and accelerate along Caldera's arc before ascending into open space.

Hers would be along soon. ■

Jeremy M. Gottwig is a former NASA librarian and programmer. He lives in Baltimore, Maryland, with his wife and young son.

ILLUSTRATION BY JACEY